



## Production of steviol glycosides in recombinant hosts

**Robertsen, Helene Lunde; Møller-Hansen, Iben; Takos, Adam Matthew; Hallwyl, Swee Chuang Lim ; Ambri, Francesca ; Quiros Asensio, Manuel; Mikkelsen, Michael Dalgaard; Houghton-Larsen, Jens; Douchin, Veronique ; Dyekjær, Jane Dannow**

*Total number of authors:*  
13

*Publication date:*  
2016

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

### *Citation (APA):*

Robertsen, H. L., Møller-Hansen, I., Takos, A. M., Hallwyl, S. C. L., Ambri, F., Quiros Asensio, M., Mikkelsen, M. D., Houghton-Larsen, J., Douchin, V., Dyekjær, J. D., Carlsen, S., Rasmussen, N. N., & Hansen, E. H. (2016). Production of steviol glycosides in recombinant hosts. (Patent No. WO2016038095).

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



- (51) International Patent Classification:  
A23L 1/236 (2006.01)
- (21) International Application Number:  
PCT/EP20 15/070620
- (22) International Filing Date:  
September 2015 (09.09.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
62/048,178 9 September 2014 (09.09.2014) US  
62/103,547 14 January 2015 (14.01.2015) US  
62/117,396 17 February 2015 (17.02.2015) US  
62/148,585 16 April 2015 (16.04.2015) US
- (71) Applicant: EVOLVA SA [CH/CH]; Duggingerstrasse 23,  
4153 Reinach (CH).
- (72) Inventors; and
- (71) Applicants : ROBERTSEN, Helene Lunde [DK/DK];  
Kogle Alle 6, DK-2970 Horsholm (DK). ANDERSEN,  
Iben Nordmark [DK/DK]; Krogholmgardsvej 93, DK-  
2950 Vedbaek (DK). TAKOS, Adam Matthew [AU/DK];  
Lyshojgardsvej 85, 3 mf, DK-2500 Valby (DK). HALL-  
WYL, Swee Chuang Lim [MY/DK]; Amalieparken 87, 4-  
4, DK-2665 Vallengbaek Strand (DK). AMBRI,  
Francesca [IT/DK]; Brodeskovparken 57, DK-3400  
Hillerod (DK). ASENSIO, Manuel Quiros [ES/DK];  
Gyngemose Parkvej 2A, st tv, DK-2860 Soborg (DK).
- (72) Inventors: MIKKELSEN, Michael Dalgaard; Bregnevej  
1, DK-3500 Vaerlose (DK). HOUGHTON-LARSEN,  
Jens; Grondalsvej 13, DK-3460 Birkerod (DK).  
DOUCHIN, Veronique; Aurikelvej 2-4th, DK-2000 Fre-  
deriksberg (DK). DYEKJAER, Jane Dannow; J.A.

Schwartz Gade 34, DK-2100 Copenhagen (DK).  
**CARLSEN, Simon**; Heimdalsgade 15, 2nd, DK-2200  
Copenhagen (DK). **RASMUSSEN, Nina Nicoline**; Glim-  
vej 29, DK-2650 Hvidovre (DK). **HANSEN, Esben Halk-  
jaer**; Howitzvej 20B 2, DK-2000 Frederiksberg (DK).

(74) Agent: SMAGGASGALE, Gillian Helen; 55 Drury Lane,  
London WC2B 5SQ (GB).

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,  
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,  
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,  
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,  
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,  
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,  
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,  
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,  
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

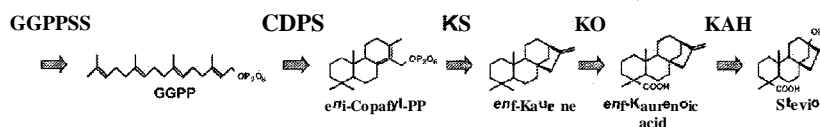
(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,  
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,  
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,  
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,  
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,  
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

- without international search report and to be republished  
upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: PRODUCTION OF STEVIOL GLYCOSIDES IN RECOMBINANT HOSTS

Figure 1



(57) Abstract: The invention relates to recombinant microorganisms and methods for producing steviol glycosides and steviol glycoside precursors.

## PRODUCTION OF STEVIOL GLYCOSIDES IN RECOMBINANT HOSTS

### BACKGROUND OF THE INVENTION

#### Field of the Invention

**[0001]** This disclosure relates to recombinant production of steviol glycosides and steviol glycoside precursors in recombinant hosts. In particular, this disclosure relates to production of steviol glycosides comprising steviol-1 3-O-glucoside (13-SMG), steviol-1,2-bioside, steviol-1,3-bioside, steviol-1 9-O-glucoside (19-SMG), stevioside, 1,3-stevioside, rubusoside, Rebaudioside A (RebA), Rebaudioside B (RebB), Rebaudioside C (RebC), Rebaudioside D (RebD), Rebaudioside E (RebE), Rebaudioside F (RebF), Rebaudioside M (RebM), Rebaudioside Q (RebQ), Rebaudioside I (RebI), dulcoside A, or isomers thereof in recombinant hosts.

#### Description of Related Art

**[0001]** Sweeteners are well known as ingredients used most commonly in the food, beverage, or confectionary industries. The sweetener can either be incorporated into a final food product during production or for stand-alone use, when appropriately diluted, as a tabletop sweetener or an at-home replacement for sugars in baking. Sweeteners include natural sweeteners such as sucrose, high fructose corn syrup, molasses, maple syrup, and honey and artificial sweeteners such as aspartame, saccharine, and sucralose. Stevia extract is a natural sweetener that can be isolated and extracted from a perennial shrub, *Stevia rebaudiana*. Stevia is commonly grown in South America and Asia for commercial production of stevia extract. Stevia extract, purified to various degrees, is used commercially as a high intensity sweetener in foods and in blends or alone as a tabletop sweetener.

**[0002]** Chemical structures for several steviol glycosides are shown in Figure 1, including the diterpene steviol and various steviol glycosides. Extracts of the Stevia plant generally comprise steviol glycosides that contribute to the sweet flavor, although the amount of each steviol glycoside often varies, *inter alia*, among different production batches.

**[0002]** As recovery and purification of steviol glycosides from the Stevia plant have proven to be labor intensive and inefficient, there remains a need for a recombinant production system that can accumulate high yields of desired steviol glycosides, such as RebD and RebM. There

also remains a need for improved production of steviol glycosides in recombinant hosts for commercial uses.

### SUMMARY OF THE INVENTION

**[0003]** it is against the above background that the present invention provides certain advantages and advancements over the prior art.

**[0004]** Although this invention disclosed herein is not limited to specific advantages or functionalities, the invention provides a recombinant host comprising one or more of:

- (a) a gene encoding an ent-kaurene oxidase (KO) polypeptide;
- (b) a gene encoding a cytochrome P450 reductase (CPR) polypeptide; and/or
- (c) a gene encoding an ent-kaurenoic acid hydroxylase (KAH) polypeptide;

wherein at least one of the genes is a recombinant gene; and

wherein the recombinant host is capable of producing a steviol glycoside precursor.

**[0005]** The invention also provides a recombinant host comprising:

- (a) a gene encoding a geranylgeranyl diphosphate synthase (GGPPS) polypeptide;
- (b) a gene encoding an ent-copalyl diphosphate synthase (CDPS) polypeptide;
- (c) a gene encoding an ent-kaurene synthase (KS) polypeptide
- (d) a gene encoding an ent-kaurene oxidase (KO) polypeptide;
- (e) a gene encoding a cytochrome P450 reductase (CPR) polypeptide; and
- (f) a gene encoding an ent-kaurenoic acid hydroxylase (KAH) polypeptide;

wherein at least one of the genes is a recombinant gene; and

wherein the recombinant host is capable of producing stevioi.

**[0006]** In one aspect of the recombinant hosts disclosed herein,

- (a) the KO polypeptide comprises a KO polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:72 or SEQ ID NO:75; 65% identity to an amino acid sequence set forth in SEQ ID NO:54; at least 70% identity to an amino acid sequence set forth in SEQ ID NO: 70, SEQ ID NO:71 , or SEQ ID NO:79; at least 40% identity to an amino acid sequence set forth in SEQ



ID NO:77; or at least 50% identity to an amino acid sequence set forth in SEQ ID NO:78;

- (b) the CPR polypeptide comprises a CPR polypeptide having at least 70% identity to an amino acid sequences set forth in SEQ ID NO:69, SEQ ID NO:74, SEQ ID NO:76, or SEQ ID NO:87; at least 80% identity to an amino acid sequence set forth in SEQ ID NO:73; at least 85% identity to an amino acid sequence set forth in SEQ ID NO:22; at least 65% identity to an amino acid sequence set forth in SEQ ID NO:28; or at least 50% identity to an amino acid sequence set forth in SEQ ID NO:98; and/or
- (c) the KAH polypeptide comprises a KAH polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:82; at least 50% identity to an amino acid sequence set forth in SEQ ID NO:91; or at least 60% identity to an amino acid sequence set forth in SEQ ID NO:68.

**[0007]** The invention further provides a recombinant host comprising one or more of:

- (a) a gene encoding a KO polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:75;
- (b) a gene encoding a KAH polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:82; and/or
- (c) a gene encoding a CPR polypeptide having at least 50% identity to an amino acid sequence set forth in SEQ ID NO:98;

wherein at least one of the genes is a recombinant gene; and

wherein the recombinant host is capable of producing a steviol glycoside precursor.

**[0008]** The invention further provides a recombinant host comprising one or more of:

- (a) a gene encoding a KO polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:70;
- (b) a gene encoding a KAH polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:82; and/or
- (c) a gene encoding a CPR polypeptide having at least 50% identity to an amino acid sequence set forth in SEQ ID NO:98;

wherein at least one of the genes is a recombinant gene; and

wherein the recombinant host is capable of producing a steviol glycoside precursor.

**[0009]** In one aspect of the recombinant hosts disclosed herein, the host further comprises a gene encoding a KO polypeptide having at least 65% identity to an amino acid sequence set forth in SEQ ID NO:54.

**[0010]** In another aspect of the recombinant hosts disclosed herein, the recombinant host further comprises a gene encoding a KAH polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:68.

**[0011]** In another aspect of the recombinant hosts disclosed herein, the recombinant host further comprises a gene encoding a KO polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:79,

**[0012]** In one aspect of the recombinant hosts disclosed herein, the host further comprises one or more of:

- (a) a gene encoding a geranylgeranyl diphosphate synthase (GGPPS) polypeptide;
  - (b) a gene encoding an ent-copalyl diphosphate synthase (CDPS) polypeptide; and/or
  - (c) a gene encoding an ent-kaurene synthase (KS) polypeptide;
- wherein at least one of the genes is a recombinant gene; and

wherein the recombinant host is capable of producing a steviol glycoside precursor.

**[0013]** In some aspects of the recombinant hosts disclosed herein,

- (a) the GGPPS polypeptide comprises a polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:49;
- (b) the CDPS polypeptide comprises a polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:37; and/or
- (c) the KS polypeptide comprises a polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:6.

**[0014]** In one aspect of the recombinant hosts disclosed herein, the recombinant host further comprises a gene encoding an endoplasmic reticulum membrane polypeptide.

**[0015]** In another aspect of the recombinant hosts disclosed herein, the endoplasmic reticulum membrane polypeptide comprises an Inheritance of cortical ER protein 2 (ICE2)

polypeptide having at least 50% identity to the amino acid sequence set forth in SEQ ID NO:1 14.

**[0016]** In one aspect of the recombinant host disclosed herein, the KO polypeptide is a fusion construct.

**[0017]** In another aspect, the fusion construct comprises a polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:118 or SEQ ID NO:120.

**[0018]** In another aspect, the fusion construct has at least 50% identity to an amino acid sequence set forth in SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:106, SEQ ID NO:108, SEQ ID NO:1 10, or SEQ ID NO:1 12.

**[0019]** in one aspect of the recombinant hosts disclosed herein, the host further comprises one or more of:

- (a) a gene encoding a UGT85C polypeptide;
- (b) a gene encoding a UGT76G polypeptide;
- (c) a gene encoding a UGT74G1 polypeptide;
- (d) a gene encoding a UGT91 D2 functional homolog polypeptide; and/or
- (e) a gene encoding an EUGT11 polypeptide;

wherein at least one of the genes is a recombinant gene; and

wherein the host is capable of producing a steviol glycoside.

**[0020]** In some aspects of the recombinant hosts disclosed herein,

- (a) the UGT85C2 polypeptide comprises a polypeptide having at least 55% identity to an amino acid sequence set forth in SEQ ID NO:30;
- (b) the UGT76G1 polypeptide comprises a polypeptide having at least 50% identity to an amino acid sequence set forth in SEQ ID NO:83;
- (c) the UGT74G1 polypeptide comprises a polypeptide having at least 55% identity to an amino acid sequence set forth in SEQ ID NO:29;
- (d) the UGT91D2 functional homolog polypeptide comprises a UGT91D2 polypeptide having 90% or greater identity to the amino acid sequence set forth in SEQ ID NO:84 or a UGT91D2e-b polypeptide having 90% or greater identity to the amino acid sequence set forth in SEQ ID NO:88; and/or

- (e) the EUGT1 1 polypeptide comprises a polypeptide having at least 65% identity to an amino acid sequence set forth in SEQ ID NO:86.

[0021] In some aspects, the recombinant hosts disclosed herein comprise a plant cell, a mammalian cell, an insect cell, a fungal cell, or a bacterial cell.

[0022] In one aspect, the bacterial cell comprises *Escherichia* bacteria cells, for example, *Escherichia coli* cells; *Lactobacillus* bacteria cells; *Lactococcus* bacteria cells; *Cornebacterium* bacteria cells; *Acetobacter* bacteria cells; *Acinetobacter* bacteria cells; or *Pseudomonas* bacterial cells.

[0023] In one aspect, the fungal cell comprises a yeast cell.

[0024] In one aspect, the yeast cell is a cell from *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, *Candida boidinii*, *Arxula adenivorans*, *Xanthophyllomyces dendrorhous*, or *Candida albicans* species.

[0025] In one aspect, the yeast cell is a *Saccharomycete*.

[0026] In one aspect, the yeast cell is a cell from the *Saccharomyces cerevisiae* species.

[0027] The invention further provides a method of producing a steviol glycoside or a steviol glycoside precursor, comprising:

- (a) growing a recombinant host disclosed herein in a culture medium, under conditions in which any of the genes disclosed herein are expressed;

wherein the steviol glycoside or the steviol glycoside precursor is synthesized by said host; and/or

- (b) optionally quantifying the steviol glycoside or the steviol glycoside precursor; and/or

- (c) optionally isolating the steviol glycoside or the steviol glycoside precursor.

[0028] In some aspects, the steviol glycoside comprises steviol-13-O-glucoside (13-SMG), steviol-1,2-bioside, steviol-1,3-bioside, steviol-19-O-glucoside (19-SMG), stevioside, 1,3-stevioside, rubusoside, Rebaudioside A (RebA), Rebaudioside B (RebB), Rebaudioside C (RebC), Rebaudioside D (RebD), Rebaudioside E (RebE), Rebaudioside F (RebF), Rebaudioside M (RebM), Rebaudioside Q (RebQ), Rebaudioside I (Rebi), dulcoside A, di-

glycosylated steviol, tri-glycosylated steviol, tetra-glycosylated steviol, penta-glycosylated steviol, hexa-glycosylated steviol, hepta-glycosylated steviol, or isomers thereof.

**[0029]** In some aspects, the steviol glycoside or steviol glycoside precursor produced by the recombinant hosts or methods disclosed herein accumulates to a detectable concentration when cultured under said conditions.

**[0030]** In some aspects, the steviol glycoside or steviol glycoside precursor produced by the recombinant hosts or methods disclosed herein has an undetectable concentration of stevia plant-derived contaminants.

**[0031]** In some aspects, the steviol glycoside or steviol glycoside precursor produced by the recombinant hosts or methods disclosed herein has a steviol glycoside composition enriched for RebD or RebM relative to the steviol glycoside composition of a wild-type *Stevia* plant.

**[0032]** These and other features and advantages of the present invention will be more fully understood from the following detailed description taken together with the accompanying claims. It is noted that the scope of the claims is defined by the recitations therein and not by the specific discussion of features and advantages set forth in the present description.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0033]** The following detailed description of the embodiments of the present invention can be best understood when read in conjunction with the following drawings, where like structure is indicated with like reference numerals and in which:

**[0034]** Figure 1 shows a schematic of the engineered biosynthetic pathway for producing steviol in yeast from geranylgeranyl diphosphate using geranylgeranyl diphosphate synthase (GGPPS), ent-copalyl diphosphate synthase (CDPS), ent-kaurene synthase (KS), ent-kaurene oxidase (KO), and ent-kaurenoic acid hydroxylase (KAH) polypeptides.

**[0035]** Figure 2 shows representative steviol glycoside glycosylation reactions catalyzed by suitable uridine S'-diphospho (UDP) glycosyl transferases (UGT) enzymes and chemical structures for several steviol glycoside compounds.

**[0036]** Figure 3 shows Rebaudioside B (RebB) production in a steviol glycoside-producing *S. cerevisiae* strain individually expressing *S. rebaudiana* K01 (SrKOI) encoded by the nucleotide sequence set forth in SEQ ID NO:59, the KO encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:55, or the KO encoded by the nucleotide sequence

set forth in SEQ ID NO:56. RebB production was measured by liquid chromatography-mass spectrometry (LC-MS) analysis as  $\mu\text{M}/\text{OD}_{600}$  of individual cultures. See Example 3.

[0037] Figure 4 shows production of ent-kaurenoic acid in steviol glycoside-producing *S. cerevisiae* strains individually expressing SrKOI encoded by the nucleotide sequence set forth in SEQ ID NO:59, the KO encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:55, or the KO encoded by the nucleotide sequence set forth in SEQ ID NO:56, as measured by LC-MS analysis of culture samples. Ent-kaurenoic acid levels were calculated as the Area under Curve (AUC) of LC-MS peaks corresponding to ent-kaurenoic acid. See Example 3.

[0038] Figure 5 shows production of total (extracellular plus intracellular) steviol glycosides in a steviol glycoside-producing *S. cerevisiae* strain overexpressing *S. rebaudiana* KAHeI (SrKAHeI; encoded by the nucleotide sequence set forth in SEQ ID NO:18) or in a steviol glycoside-producing *S. cerevisiae* strain co-expressing SrKAHeI (encoded by the nucleotide sequence set forth in SEQ ID NO:18) and a KO encoded by the nucleotide sequences set forth in any one of SEQ ID NOs: 55-60, compared to a control strain that does not overexpress SrKAHeI or express a KO encoded by the nucleotide sequence set forth in any one of SEQ ID NOs: 55-60. Production of total steviol glycosides was quantified by comparison to a standard curve. Values plotted on the y-axis in  $\mu\text{M}$  are an average of three biological replicates. See Example 4.

[0039] Figure 6 shows production of Rebaudioside A (RebA), Rebaudioside D (RebD), and Rebaudioside M (RebM) in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHeI (encoded by the nucleotide sequence set forth in SEQ ID NO:18) and further expressing either the KO encoded by the nucleotide sequence set forth in SEQ ID NO:56 or the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65. Production of RebA + RebD + RebM was measured in  $\mu\text{M}$ . See Example 4.

[0040] Figure 7 shows production of glycosylated ent-kaurenoic acid in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHeI (encoded by the nucleotide sequence set forth in SEQ ID NO:18) or in a steviol glycoside-producing strain coexpressing SrKAHeI (encoded by the nucleotide sequence set forth in SEQ ID NO:18) and a KO encoded by the nucleotide sequences set forth in any one of SEQ ID NOs: 55-60). Values were calculated as the AUC of LC-MS peaks corresponding to glycosylated ent-kaurenoic acid and as an average of three biological replicates. See Example 4.

**[0041]** Figure 8 shows production of glycosylated ent-kaurenol in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHel (encoded by the nucleotide sequence set forth in SEQ ID NO:18) or in a steviol glycoside-producing *S. cerevisiae* strain co-expressing SrKAHel (encoded by the nucleotide sequence set forth in SEQ ID NO:18) and a KO encoded by the nucleotide sequence set forth in SEQ ID NOs: 55-60). Values plotted on the y-axis were calculated as the AUC of LC-MS peaks corresponding to glycosylated ent-kaurenol. See Example 4.

**[0042]** Figure 9 shows Rebaudioside <sub>IVf</sub> (RebM) production in a steviol glycoside-producing *S. cerevisiae* strain expressing CPR1 (encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:61) or CPR7 (encoded by the nucleotide sequence set forth in SEQ ID NO:23). Values plotted on the y-axis were measured in  $\mu$ M. See Example 5.

**[0043]** Figure 10 shows Rebaudioside M (RebM) production in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHel (encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:18) and further expressing CPR4497 encoded by the nucleotide sequence set forth in SEQ ID NO:62. Values plotted on the y-axis indicate  $\mu$ M concentration of RebM. See Example 5.

**[0044]** Figure 11A shows an LC-MS chromatogram of a steviol-1 3-O-glucoside (13-SMG) standard. Figure 11B shows production of 13-SMG by a steviol glycoside-producing *S. cerevisiae* strain expressing the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80 (amino acid sequence set forth in SEQ ID NO:82). See Example 7.

**[0045]** Figure 12 shows steviol-1 3-O-glucoside (13-SMG) and Rebaudioside B (RebB) production in a steviol glycoside-producing *S. cerevisiae* strain co-expressing a KO and a CPR. The KO was selected from SrKOI (encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:59), the KO encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:63, or the KO encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:64. The cytochrome P450 reductase (CPR) polypeptide was selected from the CPR encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:66 or the CPR encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:67. Values displayed on the y-axis are  $\mu$ M concentrations of the indicated steviol glycosides. See Example 6.

**[0046]** Figure 13 shows production of steviol-1 3-O-glucoside (13-SMG) and rubusoside in a steviol glycoside-producing *S. cerevisiae* strain expressing SrKAHel (encoded by the

nucleotide sequence set forth in SEQ ID NO: 18), the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80, or the KAH encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:81. Values displayed in the y-axis are  $\mu\text{M}$  concentrations of 13-SMG and rubusoside, averaged over eight biological replicates and normalized to  $\text{OD}_{600}$  measured using a plate reader. Error bars are  $\pm$  the respective standard deviation. See Example 7.

**[0047]** Figure 14 shows cytochrome P450 reductase (CPR) polypeptide activity on cytochrome c upon incubation with microsomal protein prepared from *S. cerevisiae* strains expressing SrKAH1 (encoded by the nucleotide sequence set forth in SEQ ID NO:18) alone or in combination with CPR1 (encoded by the nucleotide sequence set forth in SEQ ID NO:61) or CPR12 (encoded by the nucleotide sequence set forth in SEQ ID NO:97). Results are shown in U/mg as an average of two biological replicates. See Example 9.

**[0048]** Figure 15A shows steviol accumulation upon 30 min incubation of ent-kaurenoic acid with microsomal protein prepared from *S. cerevisiae* strains expressing SrKAH1 (encoded by the nucleotide sequence set forth in SEQ ID NO:18) alone or in combination with CPR1 (encoded by the nucleotide sequence set forth in SEQ ID NO:61) or CPR12 (encoded by the nucleotide sequence set forth in SEQ ID NO:97). Results are shown in AUC as an average of three biological replicates. Control reactions comprised the microsomal protein described above, but these were not incubated for 30 min prior to measurement of steviol accumulation. Figure 15B shows levels of ent-kaurenoic acid following 30 min incubation of ent-kaurenoic acid with microsomal protein prepared from *S. cerevisiae* strains expressing SrKAH1 (encoded by the nucleotide sequence set forth in SEQ ID NO:18) alone or in combination with CPR1 (encoded by the nucleotide sequence set forth in SEQ ID NO:61) or CPR12 (encoded by the nucleotide sequence set forth in SEQ ID NO:97). Results are shown in  $\mu\text{M}$  as an average of three biological replicates. Control reactions comprised the microsomal protein described above but were not incubated for 30 min prior to measurement of ent-kaurenoic acid levels. See Example 9.

**[0049]** Figure 16 shows steviol-13-O-glucoside (13-SMG), 1,2-bioside, Rebaudioside B (RebB), ent-kaurenoic acid, and ent-kaurene levels accumulated by a steviol glycoside-producing *S. cerevisiae* strain expressing SrKOi (SEQ ID NO:59, SEQ ID NO:79), a KO encoded by the nucleotide sequence set forth in SEQ ID NO:65, or a fusion construct between either SrKOi or the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and the NADPH-dependent P450 oxidoreductase domain of CYP102A1 (referred to herein as the "BMR domain"). Figure 16A shows levels of 13-SMG, 1,2-bioside, and RebB measured by LC-MS for



a steviol glycoside-producing *S. cerevisiae* strain expressing SrKOl (SEQ ID NO:59, SEQ ID NO:79), a fusion construct of SrKOl and BMR (SEQ ID NO:99, SEQ ID NO:100), a fusion construct of SrKOl and BMR W1046A (SEQ ID NO:101, SEQ ID NO:102), a fusion construct of truncated SrKOl and BMR (SEQ ID NO:103, SEQ ID NO:104), a fusion construct of truncated SrKOl and BMR W1046A (SEQ ID NO:105, SEQ ID NO:106), or a control plasmid. Figure 16B shows levels of ent-kaurenoic acid and ent-kaurene measured by LC-UV for a steviol glycoside-producing *S. cerevisiae* strain expressing SrKOl (SEQ ID NO:59, SEQ ID NO:79), a fusion construct of SrKOl and BMR (SEQ ID NO:99, SEQ ID NO:100), a fusion construct of SrKOl and BMR W1046A (SEQ ID NO:101, SEQ ID NO:102), a fusion construct of truncated SrKOl and BMR (SEQ ID NO:103, SEQ ID NO:104), a fusion construct of truncated SrKOl and BMR W1046A (SEQ ID NO:105, SEQ ID NO:106), or a control plasmid. Figure 16C shows levels of 13-SMG, 1,2-bioside, and RebB measured by LC-MS for a steviol glycoside-producing *S. cerevisiae* strain expressing the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65, a fusion construct of the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and BMR (SEQ ID NO:107, SEQ ID NO:108), a fusion construct of the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and BMR W1046A (SEQ ID NO:109, SEQ ID NO:110), a fusion construct of a truncated KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and BMR W1046A (SEQ ID NO:111, SEQ ID NO:112), or a plasmid control. Figure 16D shows levels of ent-kaurenoic acid or ent-kaurene accumulated by a steviol glycoside-producing *S. cerevisiae* strain expressing the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65, a fusion construct of the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and BMR (SEQ ID NO:107, SEQ ID NO:108), a fusion construct of the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and BMR W1046A (SEQ ID NO:109, SEQ ID NO:110), a fusion construct of a truncated KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and BMR W1046A (SEQ ID NO:111, SEQ ID NO:112), or a plasmid control. See Example 10.

## DETAILED DESCRIPTION OF THE INVENTION

**[0050]** Before describing the present invention in detail, a number of terms will be defined. As used herein, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. For example, reference to a "nucleic acid" means one or more nucleic acids.

[0051] It is noted that terms like "preferably," "commonly," and "typically" are not utilized herein to limit the scope of the claimed invention or to imply that certain features are critical, essential, or even important to the structure or function of the claimed invention. Rather, these terms are merely intended to highlight alternative or additional features that can or cannot be utilized in a particular embodiment of the present invention.

[0052] For the purposes of describing and defining the present invention it is noted that the term "substantially" is utilized herein to represent the inherent degree of uncertainty that can be attributed to any quantitative comparison, value, measurement, or other representation. The term "substantially" is also utilized herein to represent the degree by which a quantitative representation can vary from a stated reference without resulting in a change in the basic function of the subject matter at issue.

[0053] Methods well known to those skilled in the art can be used to construct genetic expression constructs and recombinant cells according to this invention. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, *in vivo* recombination techniques, and polymerase chain reaction (PCR) techniques. See, for example, techniques as described in Green & Sambrook, 2012, MOLECULAR CLONING: A LABORATORY MANUAL, Fourth Edition, Cold Spring Harbor Laboratory, New York; Ausubei *et al.*, 1989, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing Associates and Wiley Interscience, New York, and PCR Protocols: A Guide to Methods and Applications (Innis *et al.*, 1990, Academic Press, San Diego, CA).

[0054] As used herein, the terms "polynucleotide", "nucleotide", "oligonucleotide", and "nucleic acid" can be used interchangeably to refer to nucleic acid comprising DNA, RNA, derivatives thereof, or combinations thereof.

[0055] As used herein, the terms "microorganism," "microorganism host," "microorganism host cell," "recombinant host," and "recombinant host cell" can be used interchangeably. As used herein, the term "recombinant host" is intended to refer to a host, the genome of which has been augmented by at least one DNA sequence. Such DNA sequences include but are not limited to genes that are not naturally present, DNA sequences that are not normally transcribed into RNA or translated into a protein ("expressed"), and other genes or DNA sequences which one desires to introduce into a host. It will be appreciated that typically the genome of a recombinant host described herein is augmented through stable introduction of one or more recombinant genes. Generally, introduced DNA is not originally resident in the host that is the recipient of the DNA, but it is within the scope of this disclosure to isolate a DNA segment from

a given host, and to subsequently introduce one or more additional copies of that DNA into the same host, e.g., to enhance production of the product of a gene or alter the expression pattern of a gene. In some instances, the introduced DNA will modify or even replace an endogenous gene or DNA sequence by, e.g., homologous recombination or site-directed mutagenesis. Suitable recombinant hosts include microorganisms.

[0056] As used herein, the term "recombinant gene" refers to a gene or DNA sequence that is introduced into a recipient host, regardless of whether the same or a similar gene or DNA sequence may already be present in such a host. "Introduced," or "augmented" in this context, is known in the art to mean introduced or augmented by the hand of man. Thus, a recombinant gene can be a DNA sequence from another species or can be a DNA sequence that originated from or is present in the same species but has been incorporated into a host by recombinant methods to form a recombinant host. It will be appreciated that a recombinant gene that is introduced into a host can be identical to a DNA sequence that is normally present in the host being transformed, and is introduced to provide one or more additional copies of the DNA to thereby permit overexpression or modified expression of the gene product of that DNA. In some aspects, said recombinant genes are encoded by cDNA. In other embodiments, recombinant genes are synthetic and/or codon-optimized for expression in *S. cerevisiae*.

[0057] As used herein, the term "engineered biosynthetic pathway" refers to a biosynthetic pathway that occurs in a recombinant host, as described herein. In some aspects, one or more steps of the biosynthetic pathway do not naturally occur in an unmodified host. In some embodiments, a heterologous version of a gene is introduced into a host that comprises an endogenous version of the gene.

[0058] As used herein, the term "endogenous" gene refers to a gene that originates from and is produced or synthesized within a particular organism, tissue, or cell. In some embodiments, the endogenous gene is a yeast gene. In some embodiments, the gene is endogenous to *S. cerevisiae*, including, but not limited to *S. cerevisiae* strain S288C. In some embodiments, an endogenous yeast gene is overexpressed. As used herein, the term "overexpress" is used to refer to the expression of a gene in an organism at levels higher than the level of gene expression in a wild type organism. See, e.g., Prelich, 2012, *Genetics* 190:841-54. In some embodiments, an endogenous yeast gene is deleted. See, e.g., Giaever & Nislow, 2014, *Genetics* 197(2):451-65. As used herein, the terms "deletion," "deleted," "knockout," and "knocked out" can be used interchangeably to refer to an endogenous gene that

has been manipulated to no longer be expressed in an organism, including, but not limited to, *S. cerevisiae*.

[0059] As used herein, the terms "heterologous sequence" and "heterologous coding sequence" are used to describe a sequence derived from a species other than the recombinant host, in some embodiments, the recombinant host is an *S. cerevisiae* cell, and a heterologous sequence is derived from an organism other than *S. cerevisiae*. A heterologous coding sequence, for example, can be from a prokaryotic microorganism, a eukaryotic microorganism, a plant, an animal, an insect, or a fungus different than the recombinant host expressing the heterologous sequence. In some embodiments, a coding sequence is a sequence that is native to the host.

[0060] A "selectable marker" can be one of any number of genes that complement host cell auxotrophy, provide antibiotic resistance, or result in a color change. Linearized DNA fragments of the gene replacement vector then are introduced into the cells using methods well known in the art (see below). Integration of the linear fragments into the genome and the disruption of the gene can be determined based on the selection marker and can be verified by, for example, PGR or Southern blot analysis. Subsequent to its use in selection, a selectable marker can be removed from the genome of the host cell by, e.g., Cre-LoxP systems (see, e.g., Gossen *et al.*, 2002, *Ann. Rev. Genetics* 36:153-173 and U.S. 2006/0014264). Alternatively, a gene replacement vector can be constructed in such a way as to include a portion of the gene to be disrupted, where the portion is devoid of any endogenous gene promoter sequence and encodes none, or an inactive fragment of, the coding sequence of the gene.

[0061] As used herein, the terms "variant" and "mutant" are used to describe a protein sequence that has been modified at one or more amino acids, compared to the wild-type sequence of a particular protein.

[0062] As used herein, the term "inactive fragment" is a fragment of the gene that encodes a protein having, e.g., less than about 10% (e.g., less than about 9%, less than about 8%, less than about 7%, less than about 6%, less than about 5%, less than about 4%, less than about 3%, less than about 2%, less than about 1%, or 0%) of the activity of the protein produced from the full-length coding sequence of the gene. Such a portion of a gene is inserted in a vector in such a way that no known promoter sequence is operably linked to the gene sequence, but that a stop codon and a transcription termination sequence are operably linked to the portion of the gene sequence. This vector can be subsequently linearized in the portion of the gene sequence

and transformed into a cell. By way of single homologous recombination, this linearized vector is then integrated in the endogenous counterpart of the gene with inactivation thereof.

**[0063]** As used herein, the term "steviol glycoside" refers to Rebaudioside A (RebA) (CAS # 58543-16-1), Rebaudioside B (RebB) (CAS # 58543-17-2), Rebaudioside C (RebC) (CAS # 63550-99-2), Rebaudioside D (RebD) (CAS # 63279-13-0), Rebaudioside E (RebE) (CAS # 63279-14-1), Rebaudioside F (RebF) (CAS # 438045-89-7), Rebaudioside M (RebM) (CAS # 1220616-44-3), Rubusoside (CAS # 63849-39-4), Dulcoside A (CAS # 64432-06-0), Rebaudioside I (RebI) (MassBank Record: FU000332), Rebaudioside Q (RebQ), 1,2-Stevioside (CAS # 57817-89-7), 1,3-Stevioside (RebG), 1,2-bioside (MassBank Record: FU000299), 1,3-bioside, Steviol-13-O-glucoside (13-SMG), Steviol-19-O-glucoside (19-SMG), a tri-glucosylated steviol glycoside, a tetra-glucosylated steviol glycoside, a penta-glucosylated steviol glycoside, a hexa-glucosylated steviol glycoside, a hepta-glucosylated steviol glycoside, and isomers thereof. See Figure 2; see also, Steviol Glycosides Chemical and Technical Assessment 69th JECFA, 2007, prepared by Harriet Wallin, Food Agric. Org.

**[0064]** As used herein, the terms "steviol glycoside precursor" and "steviol glycoside precursor compound" are used to refer to intermediate compounds in the steviol glycoside biosynthetic pathway. Steviol glycoside precursors include, but are not limited to, geranylgeranyl diphosphate (GGPP), ent-copalyl-diphosphate, ent-kaurene, ent-kaurenol, ent-kaurenal, ent-kaurenoic acid, and steviol. See Figure 1. In some embodiments, steviol glycoside precursors are themselves steviol glycoside compounds. For example, 19-SMG, rubusoside, stevioside, and RebE are steviol glycoside precursors of RebM. See Figure 2. Steviol glycosides and/or steviol glycoside precursors can be produced *in vivo* (i.e., in a recombinant host), *in vitro* (i.e., enzymatically), or by whole cell bioconversion. As used herein, the terms "produce" and "accumulate" can be used interchangeably to describe synthesis of steviol glycosides and steviol glycoside precursors *in vivo*, *in vitro*, or by whole cell bioconversion.

**[0065]** As used herein, the term "di-glycosylated steviol" can be used to refer to a steviol molecule comprising two sugar moieties, such as glucose or N-acetylglucosamine (GlcNAc). Non-limiting examples of di-glycosylated steviol molecules include steviol-1,3-bioside, steviol-1,2-bioside, rubusoside, a steviol molecule comprising two glucose moieties, a steviol molecule comprising one glucose moiety and one GlcNAc moiety, and isomers thereof.

**[0066]** As used herein, the term "tri-glycosylated steviol" can be used to refer to a steviol molecule comprising three sugar moieties, such as glucose or GlcNAc. Non-limiting examples

of tri-glycosylated steviol molecules include RebB, RebG, stevioside, a steviol molecule comprising two glucose moieties and one GlcNAc moiety, and isomers thereof.

[0067] As used herein, the term "tetra-glycosylated steviol" can be used to refer to a steviol molecule comprising four sugar moieties, such as glucose or GlcNAc. Non-limiting examples of tetra-glycosylated steviol molecules include RebA, RebE, RebQ, a steviol molecule comprising four glucose moieties, a steviol molecule comprising three glucose moieties and one GlcNAc moiety, and isomers thereof.

[0068] As used herein, the term "penta-glycosylated steviol" can be used to refer to a steviol molecule comprising five sugar moieties, such as glucose or GlcNAc. Non-limiting examples of penta-glycosylated steviol molecules include RebD, a steviol molecule comprising five glucose moieties, a steviol molecule comprising four glucose moieties and one GlcNAc moiety, and isomers thereof.

[0069] As used herein, the term "hexa-glycosylated steviol" can be used to refer to a steviol molecule comprising six sugar moieties, such as glucose or GlcNAc. Non-limiting examples of hexa-glycosylated steviol molecules include RebM, a steviol molecule comprising six glucose moieties, a steviol molecule comprising five glucose moieties and one GlcNAc moiety, and isomers thereof.

[0070] As used herein, the term "hepta-glycosylated steviol" can be used to refer to a steviol molecule comprising seven sugar moieties, such as glucose or GlcNAc. Non-limiting examples of hepta-glycosylated steviol molecules include a steviol molecule comprising seven glucose moieties and isomers thereof.

[0071] As used herein, the term "glycosylated ent-kaurenoic acid" can be used to refer to an ent-kaurenoic acid molecule comprising sugar moieties, such as glucose or GlcNAc. Non-limiting examples of glycosylated ent-kaurenoic acid molecules include ent-kaurenoic acid molecule comprising two glucose moieties and one GlcNAc moiety, an ent-kaurenoic acid molecule comprising three glucose moieties, an ent-kaurenoic acid molecule comprising one glucose moiety and one GlcNAc moiety, an ent-kaurenoic acid molecule comprising two glucose moieties, and isomers thereof.

[0072] As used herein, the term "glycosylated ent-kaurenol" can be used to refer to an ent-kaurenol molecule comprising sugar moieties, such as glucose or GlcNAc. Non-limiting examples of glycosylated ent-kaurenol molecules include an ent-kaurenol molecule comprising three glucose moieties, an ent-kaurenol molecule comprising one glucose moiety and one

GlcNAc moiety, an ent-kaureno! molecule comprising two glucose moieties, and isomers thereof.

[0073] Recombinant steviol glycoside-producing *Saccharomyces cerevisiae* (*S. cerevisiae*) strains are described in WO 2011/153378, WO 2013/022989, WO 2014/122227, and WO 2014/122328. Methods of producing steviol glycosides in recombinant hosts, by whole cell bio-conversion, and *in vitro* are also described in WO 2011/153378, WO 2013/022989, WO 2014/122227, and WO 2014/122328.

[0074] In some embodiments, steviol glycosides and/or steviol glycoside precursors are produced *in vivo* through expression of one or more enzymes involved in the steviol glycoside biosynthetic pathway in a recombinant host. For example, a steviol-producing recombinant host expressing one or more of a gene encoding a GGPPS polypeptide, a gene encoding a CDPS polypeptide, a gene encoding a KS polypeptide, a gene encoding a KO polypeptide, a gene encoding a KAH polypeptide, a gene encoding a CPR polypeptide, and a gene encoding a UGT polypeptide can produce a steviol glycoside and/or steviol glycoside precursors *in vivo*. See, e.g., Figures 1 and 2. The skilled worker will appreciate that one or more of these genes can be endogenous to the host provided that at least one (and in some embodiments, all) of these genes is a recombinant gene introduced into the recombinant host.

[0075] In another example, a recombinant host expressing a gene encoding a GGPPS polypeptide, a gene encoding a CDPS polypeptide, a gene encoding a KS polypeptide, a gene encoding a KO polypeptide, a gene encoding a KAH polypeptide, and a gene encoding a CPR polypeptide can produce steviol *in vivo*. See, e.g., Figures 1. The skilled worker will appreciate that one or more of these genes can be endogenous to the host provided that at least one (and in some embodiments, all) of these genes is a recombinant gene introduced into the recombinant host.

[0076] In another example, a steviol-producing recombinant host expressing a gene encoding a GGPPS polypeptide, a gene encoding a CDPS polypeptide, a gene encoding a KS polypeptide, a gene encoding a KO polypeptide, a gene encoding a KAH polypeptide, a gene encoding a CPR polypeptide, and one or more of a gene encoding a UGT polypeptide can produce a steviol glycoside *in vivo*. See, e.g., Figures 1 and 2. The skilled worker will appreciate that one or more of these genes can be endogenous to the host provided that at least one (and in some embodiments, all) of these genes is a recombinant gene introduced into the recombinant host.

**[0077]** Non-limiting examples of KS polypeptides are set forth in SEQ ID NOs:1-4 and SEQ ID NO:6. Non-limiting examples of KO polypeptides are set forth in SEQ ID NOs:7-10, 54, 70-72, 75, and 77-79. Non-limiting examples of KAH polypeptides are set forth in SEQ ID NOs:13-17, 68, 82, and 91. Non-limiting examples of CPR polypeptides are set forth in SEQ ID NOs:20-22, 28, 69, 73, 74, 76, 87, and 98. Non-limiting examples of CDPS polypeptides are set forth in SEQ ID NOs:33-39. Non-limiting examples of CDPS-KS polypeptides are set forth in SEQ ID NOs:40-42. Non-limiting examples of GGPPS polypeptides are set forth in SEQ ID NOs:43-50.

**[0078]** In some embodiments, a recombinant host comprises a nucleic acid encoding a UGT85C2 polypeptide (SEQ ID NO:32), a nucleic acid encoding a UGT76G1 polypeptide (SEQ ID NO:83), a nucleic acid encoding a UGT74G1 polypeptide (SEQ ID NO:29), a nucleic acid encoding a UGT91D2 polypeptide, and/or a nucleic acid encoding a EUGT1 1 polypeptide (SEQ ID NO:86). In some aspects, the UGT91D2 polypeptide can be a UGT91D2e polypeptide (SEQ ID NO:84) or a UGT91D2e-b polypeptide (SEQ ID NO:88). The skilled worker will appreciate that expression of these genes may be necessary to produce a particular steviol glycoside but that one or more of these genes can be endogenous to the host provided that at least one (and in some embodiments, all) of these genes is a recombinant gene introduced into the recombinant host. In a particular embodiment, a steviol-producing recombinant microorganism comprises exogenous nucleic acids encoding UGT85C2, UGT76G1, or UGT91D2 polypeptides. In another particular embodiment, a steviol-producing recombinant microorganism comprises exogenous nucleic acids encoding UGT85C2, UGT76G1, UGT74G1, and UGT91D2 polypeptides. In yet another particular embodiment, a steviol-producing recombinant microorganism comprises exogenous nucleic acids encoding UGT85C2, UGT76G1, UGT74G1, and EUGT11 polypeptides. In yet another particular embodiment, a steviol-producing recombinant microorganism comprises the exogenous nucleic acids encoding UGT85C2, UGT76G1, UGT74G1, UGT91D2 (including *inter alia* 91D2e, 91D2m, 91D2e-b, and functional homologs thereof), and EUGT1 1 polypeptides.

**[0079]** In certain embodiments, the steviol glycoside is RebA, RebB, RebD, and/or RebM. RebA can be synthesized in a steviol-producing recombinant microorganism expressing UGT85C2, UGT76G1, UGT74G1, and UGT91D2. RebB can be synthesized in a steviol-producing recombinant microorganism expressing UGT85C2, UGT76G1, and UGT91D2. RebD can be synthesized in a steviol-producing recombinant microorganism expressing UGT85C2, UGT76G1, UGT74G1, and UGT91D2 and/or EUGT1 1. RebM can be synthesized in a steviol-



producing recombinant microorganism expressing UGT85C2, UGT76G1, UGT74G1, and UGT91 D2 and/or EUGT1 1 (see Figure 2).

[0080] In some embodiments, steviol glycosides and/or steviol glycoside precursors are produced through contact of a steviol glycoside precursor with one or more enzymes involved in the steviol glycoside pathway *in vitro*. For example, contacting steviol with a UGT polypeptide can result in production of a steviol glycoside *in vitro*. In some embodiments, a steviol glycoside precursor is produced through contact of an upstream steviol glycoside precursor with one or more enzymes involved in the steviol glycoside pathway *in vitro*. For example, contacting ent-kaurenoic acid with a KAH enzyme can result in production of steviol *in vitro*.

[0081] In some embodiments, a steviol glycoside or steviol glycoside precursor is produced by whole cell bioconversion. For whole cell bioconversion to occur, a host cell expressing one or more enzymes involved in the steviol glycoside pathway takes up and modifies a steviol glycoside precursor in the cell; following modification *in vivo*, a steviol glycoside remains in the cell and/or is excreted into the culture medium. For example, a host cell expressing a gene encoding a UGT polypeptide can take up steviol and glycosylate steviol in the cell; following glycosylation *in vivo*, a steviol glycoside can be excreted into the culture medium. In some embodiments, the cell is permeabilized to take up a substrate to be modified or to excrete a modified product.

[0082] In some embodiments, steviol, one or more steviol glycoside precursors, and/or one or more steviol glycosides are produced by co-culturing of two or more hosts. In some embodiments, one or more hosts, each expressing one or more enzymes involved in the steviol glycoside pathway, produce steviol, one or more steviol glycoside precursors, and/or one or more steviol glycosides. For example, a host comprising a GGPPS, a CDPS, a KO, a KS, a KAH, and/or a CPR and a host comprising one or more UGTs produce one or more steviol glycosides.

[0083] In some embodiments, a steviol glycoside or steviol glycoside precursor composition produced *in vivo*, *in vitro*, or by whole cell bioconversion comprises less contaminants than a stevia extract from, *inter alia*, a stevia plant. Contaminants include plant-derived compounds that contribute to off-flavors. Potential contaminants include pigments, lipids, proteins, phenolics, saccharides, spathulenol and other sesquiterpenes, labdane diterpenes, monoterpenes, decanoic acid, 8,11,14-eicosatrienoic acid, 2-methyloctadecane, pentacosane, octacosane, tetracosane, octadecanol, stigmasterol,  $\beta$ -sitosterol,  $\alpha$ -amyrin,  $\beta$ -amyrin, lupeol,  $\beta$ -

amryin acetate, pentacyclic triterpenes, centauredin, quercitin, epi-alpha-cadinol, carophyllenes and derivatives, beta-pinene, beta-sitosterol, and gibberellin.

**[0084]** As used herein, the terms "detectable amount," "detectable concentration," "measurable amount," and "measurable concentration" refer to a level of steviol glycosides measured in AUC,  $\mu\text{M}/\text{OD}_{600}$ , mg/L,  $\mu\text{M}$ , or mM. Steviol glycoside production (*i.e.*, total, supernatant, and/or intracellular steviol glycoside levels) can be detected and/or analyzed by techniques generally available to one skilled in the art, for example, but not limited to, liquid chromatography-mass spectrometry (LC-MS), thin layer chromatography (TLC), high-performance liquid chromatography (HPLC), ultraviolet-visible spectroscopy/ spectrophotometry (UV-Vis), mass spectrometry (MS), and nuclear magnetic resonance spectroscopy (NMR).

**[0085]** As used herein, the term "undetectable concentration" refers to a level of a compound that is too low to be measured and/or analyzed by techniques such as TLC, HPLC, UV-Vis, MS, or NMR. In some embodiments, a compound of an "undetectable concentration" is not present in a steviol glycoside or steviol glycoside precursor composition.

**[0086]** As used herein, the terms "or" and "and/or" is utilized to describe multiple components in combination or exclusive of one another. For example, "x, y, and/or z" can refer to "x" alone, "y" alone, "z" alone, "x, y, and z," "(x and y) or z," "x or (y and z)," or "x or y or z." In some embodiments, "and/or" is used to refer to the exogenous nucleic acids that a recombinant cell comprises, wherein a recombinant cell comprises one or more exogenous nucleic acids selected from a group. In some embodiments, "and/or" is used to refer to production of steviol glycosides and/or steviol glycoside precursors. In some embodiments, "and/or" is used to refer to production of steviol glycosides, wherein one or more steviol glycosides are produced. In some embodiments, "and/or" is used to refer to production of steviol glycosides, wherein one or more steviol glycosides are produced through one or more of the following steps: culturing a recombinant microorganism, synthesizing one or more steviol glycosides in a recombinant microorganism, and/or isolating one or more steviol glycosides.

**[0087]** In some embodiments, the nucleotide sequence of a nucleic acid encoding a KO polypeptide is set forth in SEQ ID NO: 55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, or SEQ ID NO:60, SEQ ID NO:63, SEQ ID NO:64, or SEQ ID NO:65. In some aspects, the nucleic acid encoding the KO polypeptide has at least 70% identity to the nucleotide sequence set forth in SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59 or SEQ ID NO:60, at least 80% identity to the nucleotide sequence set forth in SEQ ID NO:56 or SEQ ID NO:58, at least 95% identity to the nucleotide sequence set forth in SEQ ID NO:63, or at least

75% identity to the nucleotide sequence set forth in SEQ ID NO:64 or SEQ ID NO:65. In some embodiments, the amino acid sequence of a KO enzyme is set forth in SEQ ID NO:54, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:78, OR SEQ ID NO:79. In some embodiments, a host cell comprises one or more copies of one or more nucleic acids encoding a KO polypeptide.

[0088] In some embodiments, expression of a KO gene set forth in SEQ ID NO:55 or SEQ ID NO:56 in a RebB-producing *S. cerevisiae* strain results in higher production of RebB compared to expression of SrKOI (SEQ ID NO:59, SEQ ID NO:79) in a RebB-producing *S. cerevisiae* strain. See Example 3.

[0089] In some embodiments, expression of a KO gene set forth in SEQ ID NO:55, SEQ ID NO:56, or SEQ ID NO:57 in an *S. cerevisiae* strain capable of producing RebB with a functional KO results in production of ent-kaurenoic acid. See Example 3.

[0090] As used herein, the terms "ent-kaurenoic acid hydroxylase" and "steviol synthase" can be used interchangeably and be abbreviated "KAH." In some embodiments, the nucleotide sequence of a nucleic acid encoding a KAH enzyme is set forth in SEQ ID NO:18, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:90, or SEQ ID NO:96. In some aspects, the nucleic acid encoding the KAH polypeptide has at least 75% identity to a nucleotide sequence set forth in SEQ ID NO:80; or at least 70% identity to a nucleotide sequence set forth in SEQ ID NO:18, SEQ ID NO:81, SEQ ID NO:90, or SEQ ID NO:96. In some embodiments, the amino acid sequence of a KAH enzyme is set forth in SEQ ID NO:68, SEQ ID NO:82, or SEQ ID NO:91. In some embodiments, a host cell comprises one or more copies of one or more nucleic acids encoding a KAH enzyme.

[0091] In some embodiments, one or more copies of SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) are expressed in an *S. cerevisiae* strain. For example, in some embodiments, two copies of SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) are expressed in an *S. cerevisiae* strain.

[0092] In some embodiments, the nucleotide sequence of a nucleic acid encoding a KAH enzyme is set forth in SEQ ID NO:80. The nucleic acid of SEQ ID NO:80 encodes a KAH with an amino acid sequence set forth in SEQ ID NO:82. A version of SEQ ID NO:80 codon-optimized for expression in *S. cerevisiae* is set forth in SEQ ID NO:81. In some embodiments, a host cell comprises one or more copies of one or more nucleic acids encoding a KAH enzyme. See Example 7.

[0093] In some embodiments, SrKAHel (SEQ ID NO:18, SEQ ID NO:68) and either the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80 or the KAH encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:81 are co-expressed in a steviol glycoside-producing *S. cerevisiae* strain. In some embodiments, co-expression of SrKAHel (SEQ ID NO:18, SEQ ID NO:68) and either the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80 or the KAH encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:81 in a steviol glycoside-producing strain results in higher production of steviol glycosides compared to a control steviol glycoside-producing strain or a steviol glycoside producing strain overexpressing SrKAHel. See Example 7 and Table 6. In some aspects, overexpressing SrKAHel results in production of 85.5  $\mu$ M 13-SMG, expression of SrKAHel and the KAH encoded by the nucleotide set forth in SEQ ID NO:80 results in production of 153.8  $\mu$ M 13-SMG, and expression of SrKAHel and the KAH encoded by the nucleotide set forth in SEQ ID NO:81 results in production of 130.5  $\mu$ M 13-SMG.

[0094] In some embodiments, a KO gene is expressed in a steviol glycoside-producing *S. cerevisiae* strain that further overexpresses SrKAHel (SEQ ID NO:18, SEQ ID NO:68). In some embodiments, expression of a KO gene of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, or SEQ ID NO:60, SEQ ID NO:65 in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHel results in higher expression of steviol glycosides compared to a control steviol-glycoside producing strain or a steviol glycoside-producing strain overexpressing SrKAHel (SEQ ID NO:18, SEQ ID NO:68). See Example 4.

[0095] In some embodiments, expression of a KO gene of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, or SEQ ID NO:60 in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHel (SEQ ID NO:18, SEQ ID NO:68) results in higher levels of glycosylated ent-kaurenoic acid compared to a control *S. cerevisiae* strain. See Example 4.

[0096] In some embodiments, expression of a KO gene of SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:59, or SEQ ID NO:60 in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHel (SEQ ID NO:18, SEQ ID NO:68) results in improved metabolic conversion of a glycosylated ent-kaurenol intermediate compound relative to a control *S. cerevisiae* strain or a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHel (SEQ ID NO:18, SEQ ID NO:68). See Example 4.

[0097] In some embodiments, a KAH is a *Prunus* KAH, such as a *Prunus avium*, *Prunus mume*, or *Prunus persica* KAH. In some embodiments, a KAH is a KAH of the CYP72A219 or CYP71A219-like family. In some embodiments, the nucleotide sequence of a nucleic acid

encoding a KAH enzyme is set forth in SEQ ID NO:90 or SEQ ID NO:96. The nucleic acids of SEQ ID NO:90 and SEQ ID NO:96 encode a KAH from *Prunus avium* with an amino acid sequence set forth in SEQ ID NO:91. In some embodiments, a KAH polypeptide is a polypeptide with an amino acid sequence set forth in SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, or SEQ ID NO:95. In some embodiments, a KAH polypeptide is a KAH polypeptide with at least 50% sequence identity to an amino acid sequence set forth in SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, or SEQ ID NO:95. In some embodiments, expression of a gene encoding a polypeptide having at least 50% sequence identity to an amino acid sequence set forth in SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, or SEQ ID NO:95 in a recombinant host results in production of a steviol glycoside or steviol glycoside precursor, such as 13-SMG and/or rubusoside. See Example 8.

**[0098]** In some embodiments, the nucleotide sequence of the nucleic acid encoding a CPR enzyme is set forth in SEQ ID NO:23, SEQ ID NO:51, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:66, SEQ ID NO:67, or SEQ ID NO:97. In some aspects, the nucleic acid encoding the CPR polypeptide has at least 75% identity to the nucleotide sequence set forth in SEQ ID NO:23, SEQ ID NO:61, or SEQ ID NO:62, or at least 70% identity to the nucleotide sequence set forth in SEQ ID NO:24, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:51, or SEQ ID NO:97. In some embodiments, the amino acid sequence of the CPR enzyme is set forth in SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:69, SEQ ID NO:73, SEQ ID NO:74, or SEQ ID NO:76, SEQ ID NO:87, or SEQ ID NO:98. In some embodiments, a host cell comprises one or more copies of one or more nucleic acids encoding a CPR enzyme.

**[0099]** In a non-limiting example, SrKAHel is activated by the *S. cerevisiae* CPR encoded by gene NCP1 (YHR042W). Enhanced activation of the KAH encoded by SrKAHel is observed when the *Arabidopsis thaliana* CPR encoded by the gene ATR2 (SEQ ID NO:51) or the *S. rebaudiana* CPR encoded by the genes CPR7 (SEQ ID NO:23) or CPR8 (SEQ ID NO:24, SEQ ID NO:28) are co-expressed in a recombinant cell. Amino acid sequences of the *A. thaliana* polypeptides ATR1 and ATR2 are set forth in SEQ ID NO:25 and SEQ ID NO:26, respectively. The *S. rebaudiana* polypeptides CPR7 and CPR8 are set forth in SEQ ID NO:27 and SEQ ID NO:28, respectively.

**[00100]** In some embodiments, expression of CPR1 (SEQ ID NO:61, SEQ ID NO:76) or of CPR7 in the steviol glycoside-producing *S. cerevisiae* strain co-expressing *S. rebaudiana* CPR8 (SEQ ID NO:24, SEQ ID NO:28) and *A. thaliana* ATR2 (SEQ ID NO:51) results in higher levels of RebM compared to a control steviol glycoside-producing *S. cerevisiae* strain expressing *S.*

*rebaudiana* CPR8 (SEQ ID NO:24, SEQ ID NO:28) and *A. thaliana* ATR2 (SEQ ID NO:51). In some embodiments, expression of the CPR set forth in SEQ ID NO:62 in a steviol glycoside-producing *S. cerevisiae* strain overexpressing SrKAHel (SEQ ID NO:18, SEQ ID NO:68) results in higher levels of RebM compared to a steviol glycoside-producing *S. cerevisiae* strain that does not express the nucleic acid set forth in SEQ ID NO:62 or overexpress SrKAHel. See Example 5.

**[00101]** In some embodiments, co-expression of SrKOI (SEQ ID NO:59, SEQ ID NO:79) and a CPR gene of SEQ ID NO:66 or SEQ ID NO:77 in a RebB-producing strain results in higher production of 13-SMG and RebB than co-expression of a KO gene of SEQ ID NO:63 or SEQ ID NO:64 and a CPR gene of SEQ ID NO:66 or SEQ ID NO:77. See Example 6.

**[00102]** In some embodiments, CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) activates cytochrome c. In some embodiments, CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) in the presence of SrKAHel (SEQ ID NO:18, SEQ ID NO:68) activate cytochrome c. In some embodiments, CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) regulate conversion of ent-kaurenoic acid to steviol. In some embodiments, CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) in combination with SrKAHel (SEQ ID NO:18, SEQ ID NO:68) convert ent-kaurenoic acid to steviol. In some embodiments, steviol production is detected upon incubation of ent-kaurenoic acid with microsomal protein prepared from *S. cerevisiae* strains expressing CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) in combination with SrKAHel (SEQ ID NO:18, SEQ ID NO:68). In some embodiments, expression of CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) in a recombinant host results in production of a steviol glycoside or steviol glycoside precursor. See Example 9.

**[00103]** In some embodiments, a steviol glycoside-producing strain expresses a fusion construct comprising a KO and the NADPH-dependent P450 oxidoreductase domain of CYP102A1, referred to herein as "BMR." The codon-optimized nucleotide sequence encoding the BMR polypeptide is set forth in SEQ ID NO:117; the BMR amino acid sequence is set forth in SEQ ID NO:118. In some embodiments, BMR is a mutant BMR, including, but not limited to a BMR W1046A mutant (SEQ ID NO:119, SEQ ID NO:120). The BMR mutant can be specific for NADH. In some embodiments, the KO-BMR fusion construct comprises a linker (SEQ ID NO:121, SEQ ID NO:122). In some embodiments, the KO of the fusion construct is SrKOI (SEQ ID NO:59, SEQ ID NO:79) or the KO encoded by the nucleotide sequence set forth in

SEQ ID NO:65 (corresponding to the amino acid sequence set forth in SEQ ID NO:75). In some embodiments, the KO of the fusion construct is a truncated KO. Exemplary KO-BMR fusion constructs are set forth in SEQ ID NOs:99-112. See Example 10.

**[00104]** In some embodiments, expression of SrK01-BMR fusion constructs (SEQ ID NOs:99-106) in a steviol glycoside-producing strain results in an increase in ent-kaurenoic acid, 13-SMG, and RebB levels, compared to expression of SrK01 (SEQ ID NO:59, SEQ ID NO:79) in a steviol glycoside-producing strain. In some embodiments, expression of a fusion construct (SEQ ID NO:107, SEQ ID NO:108) in a steviol glycoside-producing strain results in greater conversion of ent-kaurene to ent-kaurenoic acid and greater conversion of ent-kaurenoic acid to 13-SMG, compared to expression of the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 in a steviol glycoside-producing strain. In some embodiments, expression of a fusion construct comprising the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and the W1046A mutant BMR (SEQ ID NO:109, SEQ ID NO:110) results in increased ent-kaurenoic acid levels. See Figure 16 (B and D) and Example 10.

**[00105]** In some embodiments, a steviol glycoside-producing strain comprises inheritance of cortical ER protein 2 (ICE2; SEQ ID NO:113, SEQ ID NO:114). ICE2 is also referred to as YIL090W. In some aspects, ICE2 is overexpressed. ICE2 can be expressed in a strain comprising CPR1 (SEQ ID NO:61, SEQ ID NO:76) and/or CPR12 (SEQ ID NO:97, SEQ ID NO:98). In some embodiments, a steviol glycoside-producing strain comprises two copies of ICE2. In some embodiments, expression of ICE2 increases ent-kaurene metabolism (resulting in decreased accumulation of ent-kaurene, ent-kaurenol, ent-kaurenal, and ent-kaurenol glycosides), resulting in increased accumulation of steviol glycosides, compared to a control strain. See Table 10 and Example 11.

**[00106]** In some embodiments, expression of the KO encoded by nucleotide sequence set forth in SEQ ID NO:56 in a steviol glycoside-producing strain cultivated by fermentation results in a lower accumulation of ent-kaurene compounds, compared to a control steviol glycoside-producing strain. In some aspects, higher levels of ent-kaurenoic acid and steviol glycosides result, as compared to a control strain. In some embodiments, expression of the KAH encoded by nucleotide sequence set forth in SEQ ID NO:80, the KO encoded by nucleotide sequence set forth in SEQ ID NO:56, and the KO encoded by nucleotide sequence set forth in SEQ ID NO:65 in a steviol glycoside-producing strain cultivated by fermentation results in decreased accumulation of ent-kaurene, ent-kaurenol, ent-kaurenal, ent-kaurenol glycosides, ent-kaurenoic acid, and ent-kaurenoic acid glycosides and increased production of steviol glycosides, as

compared to a control strain. In some embodiments, expression of CPR12 (SEQ ID NO:97, SEQ ID NO:98), the KAH encoded by nucleotide sequence set forth in SEQ ID NO:80, and the KO encoded by nucleotide sequence set forth in SEQ ID NO:56 cultivated by fermentation results in decreased ent-kaurene, ent-kaurenol, ent-kaurenal, ent-kaurenol glycosides, ent-kaurenoic acid, and ent-kaurenoic acid glycosides accumulation and higher levels of steviol glycosides, as compared to a control strain. See Table 12 and Example 12.

### **Functional Homologs**

**[00107]** Functional homologs of the polypeptides described above are also suitable for use in producing steviol glycosides in a recombinant host. A functional homolog is a polypeptide that has sequence similarity to a reference polypeptide, and that carries out one or more of the biochemical or physiological function(s) of the reference polypeptide. A functional homolog and the reference polypeptide can be a natural occurring polypeptide, and the sequence similarity can be due to convergent or divergent evolutionary events. As such, functional homologs are sometimes designated in the literature as homologs, or orthologs, or paralogs. Variants of a naturally occurring functional homolog, such as polypeptides encoded by mutants of a wild type coding sequence, can themselves be functional homologs. Functional homologs can also be created via site-directed mutagenesis of the coding sequence for a polypeptide, or by combining domains from the coding sequences for different naturally-occurring polypeptides ("domain swapping"). Techniques for modifying genes encoding functional polypeptides described herein are known and include, *inter alia*, directed evolution techniques, site-directed mutagenesis techniques and random mutagenesis techniques, and can be useful to increase specific activity of a polypeptide, alter substrate specificity, alter expression levels, alter subcellular location, or modify polypeptide-polypeptide interactions in a desired manner. Such modified polypeptides are considered functional homologs. The term "functional homolog" is sometimes applied to the nucleic acid that encodes a functionally homologous polypeptide.

**[00108]** Functional homologs can be identified by analysis of nucleotide and polypeptide sequence alignments. For example, performing a query on a database of nucleotide or polypeptide sequences can identify homologs of steviol glycoside biosynthesis polypeptides. Sequence analysis can involve BLAST, Reciprocal BLAST, or PSI-BLAST analysis of non-redundant databases using a KO, KAH, or CPR amino acid sequence as the reference sequence. Amino acid sequence is, in some instances, deduced from the nucleotide sequence. Those polypeptides in the database that have greater than 40% sequence identity are candidates for further evaluation for suitability as a steviol glycoside biosynthesis polypeptide.



Amino acid sequence similarity allows for conservative amino acid substitutions, such as substitution of one hydrophobic residue for another or substitution of one polar residue for another. If desired, manual inspection of such candidates can be carried out in order to narrow the number of candidates to be further evaluated. Manual inspection can be performed by selecting those candidates that appear to have domains present in steviol glycoside biosynthesis polypeptides, e.g., conserved functional domains. In some embodiments, nucleic acids and polypeptides are identified from transcriptome data based on expression levels rather than by using BLAST analysis.

[00109] Conserved regions can be identified by locating a region within the primary amino acid sequence of a steviol glycoside biosynthesis polypeptide that is a repeated sequence, forms some secondary structure (e.g., helices and beta sheets), establishes positively or negatively charged domains, or represents a protein motif or domain. See, e.g., the Pfam web site describing consensus sequences for a variety of protein motifs and domains on the World Wide Web at [sanger.ac.uk/Software/Pfam/](http://sanger.ac.uk/Software/Pfam/) and [pfam.janelia.org/](http://pfam.janelia.org/). The information included at the Pfam database is described in Sonnhammer *et al.*, *Nucl. Acids Res.*, 26:320-322 (1998); Sonnhammer *et al.*, *Proteins*, 28:405-420 (1997); and Bateman *et al.*, *Nucl. Acids Res.*, 27:260-262 (1999). Conserved regions also can be determined by aligning sequences of the same or related polypeptides from closely related species. Closely related species preferably are from the same family. In some embodiments, alignment of sequences from two different species is adequate to identify such homologs.

[00110] Typically, polypeptides that exhibit at least about 40% amino acid sequence identity are useful to identify conserved regions. Conserved regions of related polypeptides exhibit at least 45% amino acid sequence identity (e.g., at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% amino acid sequence identity). In some embodiments, a conserved region exhibits at least 92%, 94%, 96%, 98%, or 99% amino acid sequence identity.

[00111] For example, polypeptides suitable for producing steviol in a recombinant host include functional homologs of KO, KAH, and CPR.

[00112] Methods to modify the substrate specificity of, for example, KO, KAH, or CPR, are known to those skilled in the art, and include without limitation site-directed/rational mutagenesis approaches, random directed evolution approaches and combinations in which random mutagenesis/saturation techniques are performed near the active site of the enzyme. For example see Osmani *et al.*, 2009, *Phytochemistry* 70: 325-347.

**[00113]** A candidate sequence typically has a length that is from 80% to 200% of the length of the reference sequence, e.g., 82, 85, 87, 89, 90, 93, 95, 97, 99, 100, 105, 110, 115, 120, 130, 140, 150, 160, 170, 180, 190, or 200% of the length of the reference sequence. A functional homolog polypeptide typically has a length that is from 95% to 105% of the length of the reference sequence, e.g., 90, 93, 95, 97, 99, 100, 105, 110, 115, or 120% of the length of the reference sequence, or any range between. A% identity for any candidate nucleic acid or polypeptide relative to a reference nucleic acid or polypeptide can be determined as follows. A reference sequence (e.g., a nucleic acid sequence or an amino acid sequence described herein) is aligned to one or more candidate sequences using the computer program ClustalW (version 1.83, default parameters), which allows alignments of nucleic acid or polypeptide sequences to be carried out across their entire length (global alignment). Chenna *et al.*, 2003, *Nucleic Acids Res.* 31(13):3497-500.

**[00114]** ClustalW calculates the best match between a reference and one or more candidate sequences, and aligns them so that identities, similarities and differences can be determined. Gaps of one or more residues can be inserted into a reference sequence, a candidate sequence, or both, to maximize sequence alignments. For fast pairwise alignment of nucleic acid sequences, the following default parameters are used: word size: 2; window size: 4; scoring method: % age; number of top diagonals: 4; and gap penalty: 5. For multiple alignment of nucleic acid sequences, the following parameters are used: gap opening penalty: 10.0; gap extension penalty: 5.0; and weight transitions: yes. For fast pairwise alignment of protein sequences, the following parameters are used: word size: 1; window size: 5; scoring method: % age; number of top diagonals: 5; gap penalty: 3. For multiple alignment of protein sequences, the following parameters are used: weight matrix: blosum; gap opening penalty: 10.0; gap extension penalty: 0.05; hydrophilic gaps: on; hydrophilic residues: Gly, Pro, Ser, Asn, Asp, Gin, Glu, Arg, and Lys; residue-specific gap penalties: on. The ClustalW output is a sequence alignment that reflects the relationship between sequences. ClustalW can be run, for example, at the Baylor College of Medicine Search Launcher site on the World Wide Web ([searchlauncher.bcm.tmc.edu/multi-align/multi-align.html](http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html)) and at the European Bioinformatics Institute site on the World Wide Web ([ebi.ac.uk/clustalw](http://ebi.ac.uk/clustalw)).

**[00115]** To determine % identity of a candidate nucleic acid or amino acid sequence to a reference sequence, the sequences are aligned using ClustalW, the number of identical matches in the alignment is divided by the length of the reference sequence, and the result is multiplied by 100. It is noted that the % identity value can be rounded to the nearest tenth. For

example, 78.11, 78.12, 78.13, and 78.14 are rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 are rounded up to 78.2.

**[001 16]** It will be appreciated that functional KO, KAH, or CPR proteins can include additional amino acids that are not involved in the enzymatic activities carried out by the enzymes. In some embodiments, KO, KAH, or CPR proteins are fusion proteins. The terms "chimera," "fusion polypeptide," "fusion protein," "fusion enzyme," "fusion construct," "chimeric protein," "chimeric polypeptide," "chimeric construct," and "chimeric enzyme" can be used interchangeably herein to refer to proteins engineered through the joining of two or more genes that code for different proteins. In some embodiments, a nucleic acid sequence encoding a KO, KAH, or CPR polypeptide can include a tag sequence that encodes a "tag" designed to facilitate subsequent manipulation (e.g., to facilitate purification or detection), secretion, or localization of the encoded polypeptide. Tag sequences can be inserted in the nucleic acid sequence encoding the polypeptide such that the encoded tag is located at either the carboxyl or amino terminus of the polypeptide. Non-limiting examples of encoded tags include green fluorescent protein (GFP), human influenza hemagglutinin (HA), glutathione S transferase (GST), polyhistidine-tag (HIS tag), and Flag™ tag (Kodak, New Haven, CT). Other examples of tags include a chloroplast transit peptide, a mitochondrial transit peptide, an amyloplast peptide, signal peptide, or a secretion tag.

**[001 17]** In some embodiments, a fusion protein is a protein altered by domain swapping. As used herein, the term "domain swapping" is used to describe the process of replacing a domain of a first protein with a domain of a second protein. In some embodiments, the domain of the first protein and the domain of the second protein are functionally identical or functionally similar. In some embodiments, the structure and/or sequence of the domain of the second protein differs from the structure and/or sequence of the domain of the first protein. In some embodiments, a KO polypeptide is altered by domain swapping. See Example 10.

#### **Steviol and Steviol Glycoside Biosynthesis Nucleic Acids**

**[001 18]** A recombinant gene encoding a polypeptide described herein comprises the coding sequence for that polypeptide, operably linked in sense orientation to one or more regulatory regions suitable for expressing the polypeptide. Because many microorganisms are capable of expressing multiple gene products from a polycistronic mRNA, multiple polypeptides can be expressed under the control of a single regulatory region for those microorganisms, if desired. A coding sequence and a regulatory region are considered to be operably linked when the regulatory region and coding sequence are positioned so that the regulatory region is effective

for regulating transcription or translation of the sequence. Typically, the translation initiation site of the translational reading frame of the coding sequence is positioned between one and about fifty nucleotides downstream of the regulatory region for a monocistronic gene.

[00119] In many cases, the coding sequence for a polypeptide described herein is identified in a species other than the recombinant host, *i.e.*, is a heterologous nucleic acid. Thus, if the recombinant host is a microorganism, the coding sequence can be from other prokaryotic or eukaryotic microorganisms, from plants or from animals. In some case, however, the coding sequence is a sequence that is native to the host and is being reintroduced into that organism. A native sequence can often be distinguished from the naturally occurring sequence by the presence of non-natural sequences linked to the exogenous nucleic acid, *e.g.*, non-native regulatory sequences flanking a native sequence in a recombinant nucleic acid construct. In addition, stably transformed exogenous nucleic acids typically are integrated at positions other than the position where the native sequence is found. "Regulatory region" refers to a nucleic acid having nucleotide sequences that influence transcription or translation initiation and rate, and stability and/or mobility of a transcription or translation product. Regulatory regions include, without limitation, promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, protein binding sequences, 5' and 3' untranslated regions (UTRs), transcriptional start sites, termination sequences, polyadenylation sequences, introns, and combinations thereof. A regulatory region typically comprises at least a core (basal) promoter. A regulatory region also may include at least one control element, such as an enhancer sequence, an upstream element or an upstream activation region (UAR). A regulatory region is operably linked to a coding sequence by positioning the regulatory region and the coding sequence so that the regulatory region is effective for regulating transcription or translation of the sequence. For example, to operably link a coding sequence and a promoter sequence, the translation initiation site of the translational reading frame of the coding sequence is typically positioned between one and about fifty nucleotides downstream of the promoter. A regulatory region can, however, be positioned as much as about 5,000 nucleotides upstream of the translation initiation site, or about 2,000 nucleotides upstream of the transcription start site.

[00120] The choice of regulatory regions to be included depends upon several factors, including, but not limited to, efficiency, selectability, inducibility, desired expression level, and preferential expression during certain culture stages. It is a routine matter for one of skill in the art to modulate the expression of a coding sequence by appropriately selecting and positioning regulatory regions relative to the coding sequence. It will be understood that more than one

regulatory region may be present, e.g., introns, enhancers, upstream activation regions, transcription terminators, and inducible elements.

[00121] One or more genes can be combined in a recombinant nucleic acid construct in "modules" useful for a discrete aspect of steviol and/or steviol glycoside production. Combining a plurality of genes in a module, particularly a polycistronic module, facilitates the use of the module in a variety of species. For example, a steviol biosynthesis gene cluster, or a UGT gene cluster, can be combined in a polycistronic module such that, after insertion of a suitable regulatory region, the module can be introduced into a wide variety of species. As another example, a UGT gene cluster can be combined such that each UGT coding sequence is operably linked to a separate regulatory region, to form a UGT module. Such a module can be used in those species for which monocistronic expression is necessary or desirable. In addition to genes useful for steviol or steviol glycoside production, a recombinant construct typically also contains an origin of replication, and one or more selectable markers for maintenance of the construct in appropriate species.

[00122] It will be appreciated that because of the degeneracy of the genetic code, a number of nucleic acids can encode a particular polypeptide; *i.e.*, for many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid. Thus, codons in the coding sequence for a given polypeptide can be modified such that optimal expression in a particular host is obtained, using appropriate codon bias tables for that host (e.g., microorganism). As isolated nucleic acids, these modified sequences can exist as purified molecules and can be incorporated into a vector or a virus for use in constructing modules for recombinant nucleic acid constructs.

[0003] In some cases, it is desirable to inhibit one or more functions of an endogenous polypeptide in order to divert metabolic intermediates towards steviol or steviol glycoside biosynthesis. For example, it may be desirable to downregulate synthesis of sterols in a yeast strain in order to further increase steviol or steviol glycoside production, e.g., by downregulating squalene epoxidase. As another example, it may be desirable to inhibit degradative functions of certain endogenous gene products, e.g., glycohydrolases that remove glucose moieties from secondary metabolites or phosphatases as discussed herein. In such cases, a nucleic acid that overexpresses the polypeptide or gene product may be included in a recombinant construct that is transformed into the strain. Alternatively, mutagenesis can be used to generate mutants in genes for which it is desired to increase or enhance function.

### **Host Microorganisms**

[00123] Recombinant hosts can be used to express polypeptides for the producing steviol glycosides, including mammalian, insect, plant, and algal cells. A number of prokaryotes and eukaryotes are also suitable for use in constructing the recombinant microorganisms described herein, e.g., gram-negative bacteria, yeast, and fungi. A species and strain selected for use as a steviol glycoside production strain is first analyzed to determine which production genes are endogenous to the strain and which genes are not present. Genes for which an endogenous counterpart is not present in the strain are advantageously assembled in one or more recombinant constructs, which are then transformed into the strain in order to supply the missing function(s).

[00124] Typically, the recombinant microorganism is grown in a fermenter at a defined temperature(s) for a desired period of time. The constructed and genetically engineered microorganisms provided by the invention can be cultivated using conventional fermentation processes, including, *inter alia*, chemostat, batch, fed-batch cultivations, semi-continuous fermentations such as draw and fill, continuous perfusion fermentation, and continuous perfusion cell culture. Depending on the particular microorganism used in the method, other recombinant genes such as isopentenyl biosynthesis genes and terpene synthase and cyclase genes may also be present and expressed. Levels of substrates and intermediates, e.g., isopentenyl diphosphate, dimethylallyl diphosphate, GGPP, ent-kaurene and ent-kaurenoic acid, can be determined by extracting samples from culture media for analysis according to published methods.

[00125] Carbon sources of use in the instant method include any molecule that can be metabolized by the recombinant host cell to facilitate growth and/or production of the steviol glycosides. Examples of suitable carbon sources include, but are not limited to, sucrose (e.g., as found in molasses), fructose, xylose, ethanol, glycerol, glucose, cellulose, starch, cellobiose or other glucose-comprising polymer. In embodiments employing yeast as a host, for example, carbon sources such as sucrose, fructose, xylose, ethanol, glycerol, and glucose are suitable. The carbon source can be provided to the host organism throughout the cultivation period or alternatively, the organism can be grown for a period of time in the presence of another energy source, e.g., protein, and then provided with a source of carbon only during the fed-batch phase.

[00126] After the recombinant microorganism has been grown in culture for the desired period of time, steviol and/or one or more steviol glycosides can then be recovered from the culture using various techniques known in the art. In some embodiments, a permeabilizing

agent can be added to aid the feedstock entering into the host and product getting out. For example, a crude lysate of the cultured microorganism can be centrifuged to obtain a supernatant. The resulting supernatant can then be applied to a chromatography column, e.g., a C-18 column, and washed with water to remove hydrophilic compounds, followed by elution of the compound(s) of interest with a solvent such as methanol. The compound(s) can then be further purified by preparative HPLC. See also, WO 2009/140394.

[00127] It will be appreciated that the various genes and modules discussed herein can be present in two or more recombinant hosts rather than a single host. When a plurality of recombinant hosts is used, they can be grown in a mixed culture to accumulate steviol and/or steviol glycosides.

[00128] Alternatively, the two or more hosts each can be grown in a separate culture medium and the product of the first culture medium, e.g., steviol, can be introduced into second culture medium to be converted into a subsequent intermediate, or into an end product such as, for example, RebA. The product produced by the second, or final host is then recovered. It will also be appreciated that in some embodiments, a recombinant host is grown using nutrient sources other than a culture medium and utilizing a system other than a fermenter.

[00129] Exemplary prokaryotic and eukaryotic species are described in more detail below. However, it will be appreciated that other species can be suitable. For example, suitable species can be in a genus such as *Agaricus*, *Aspergillus*, *Bacillus*, *Candida*, *Corynebacterium*, *Eremothecium*, *Escherichia*, *Fusarium*/*Gibberella*, *Kluyveromyces*, *Laetiporus*, *Lentinus*, *Phaffia*, *Phanerochaete*, *Pichia*, *Physcomitrella*, *Rhodoturula*, *Saccharomyces*, *Schizosaccharomyces*, *Sphaceloma*, *Xanthophyllomyces* or *Yarrowia*. Exemplary species from such genera include *Lentinus tigrinus*, *Laetiporus sulphureus*, *Phanerochaete chrysosporium*, *Pichia pastoris*, *Cyberlindnera jadinii*, *Physcomitrella patens*, *Rhodoturula glutinis*, *Rhodoturula mucilaginosa*, *Phaffia rhodozyma*, *Xanthophyllomyces dendrorhous*, *Fusarium fujikuroi*/*Gibberella fujikuroi*, *Candida utilis*, *Candida glabrata*, *Candida albicans*, and *Yarrowia lipolytica*.

[00130] In some embodiments, a microorganism can be a prokaryote such as *Escherichia* bacteria cells, for example, *Escherichia coli* cells; *Lactobacillus* bacteria cells; *Lactococcus* bacteria cells; *Corynebacterium* bacteria cells; *Acetobacter* bacteria cells; *Acinetobacter* bacteria cells; or *Pseudomonas* bacterial cells.

[00131] In some embodiments, a microorganism can be an Ascomycete such as *Gibberella fujikuroi*, *Kluyveromyces lactis*, *Schizosaccharomyces pombe*, *Aspergillus niger*, *Yarrowia lipolytica*, *Ashbya gossypii*, or *S. cerevisiae*.

[00132] in some embodiments, a microorganism can be an algal cell such as *Blakeslea trispora*, *Dunaliella salina*, *Haematococcus pluvialis*, *Chlorella* sp., *Undaria pinnatifida*, *Sargassum*, *Laminaria japonica*, *Scenedesmus almeriensis* species.

[00133] In some embodiments, a microorganism can be a cyanobacterial cell such as *Blakeslea trispora*, *Dunaliella salina*, *Haematococcus pluvialis*, *Chlorella* sp., *Undaria pinnatifida*, *Sargassum*, *Laminaria japonica*, *Scenedesmus almeriensis*.

#### *Saccharomyces* spp.

[00134] *Saccharomyces* is a widely used chassis organism in synthetic biology, and can be used as the recombinant microorganism platform. For example, there are libraries of mutants, plasmids, detailed computer models of metabolism and other information available for *S. cerevisiae*, allowing for rational design of various modules to enhance product yield. Methods are known for making recombinant microorganisms.

#### *Aspergillus* spp.

[00135] *Aspergillus* **species** such as *A. oryzae*, *A. niger* and *A. sojae* are widely used microorganisms in food production and can also be used as the recombinant microorganism platform. Nucleotide sequences are available for genomes of *A. nidulans*, *A. fumigatus*, *A. oryzae*, *A. clavatus*, *A. flavus*, *A. niger*, and *A. terreus*, allowing rational design and modification of endogenous pathways to enhance flux and increase product yield. Metabolic models have been developed for *Aspergillus*, as well as transcriptomic studies and proteomics studies. *A. niger* is cultured for the industrial production of a number of food ingredients such as citric acid and gluconic acid, and thus species such as *A. niger* are generally suitable for producing steviol glycosides.

#### *E. coli*

[00136] *E. coli*, another widely used platform organism in synthetic biology, can also be used as the recombinant microorganism platform. Similar to *Saccharomyces*, there are libraries of mutants, plasmids, detailed computer models of metabolism and other information available for *E. coli*, allowing for rational design of various modules to enhance product yield. Methods



similar to those described above for *Saccharomyces* can be used to make recombinant *E. coli* microorganisms.

*Agaricus*, *Gibberella*, and *Phanerochaete* spp.

**[00137]** *Agaricus*, *Gibberella*, and *Phanerochaete* spp. can be useful because they are known to produce large amounts of isoprenoids in culture. Thus, the terpene precursors for producing large amounts of steviol glycosides are already produced by endogenous genes. Thus, modules comprising recombinant genes for steviol glycoside biosynthesis polypeptides can be introduced into species from such genera without the necessity of introducing mevalonate or MEP pathway genes.

*Arxuia adenivorans* (*Blastobotrys adenivorans*)

**[00138]** *Arxuia adenivorans* is dimorphic yeast (it grows as budding yeast like the baker's yeast up to a temperature of 42°C, above this threshold it grows in a filamentous form) with unusual biochemical characteristics. It can grow on a wide range of substrates and can assimilate nitrate. It has successfully been applied to the generation of strains that can produce natural plastics or the development of a biosensor for estrogens in environmental samples.

*Yarrowia lipolytica*

**[00139]** *Yarrowia lipolytica* is dimorphic yeast (see *Arxuia adenivorans*) and belongs to the family Hemiascomycetes. The entire genome of *Yarrowia lipolytica* is known. *Yarrowia* species is aerobic and considered to be non-pathogenic. *Yarrowia* is efficient in using hydrophobic substrates (e.g. alkanes, fatty acids, oils) and can grow on sugars. It has a high potential for industrial applications and is an oleaginous microorganism. *Yarrowia lipolytica* can accumulate lipid content to approximately 40% of its dry cell weight and is a model organism for lipid accumulation and remobilization. See e.g., Nicaud, 2012, *Yeast* 29(10):409-18; Beopoulos et al., 2009, *Biochimie* 91(6):692-6; Bankar et al., 2009, *Appl Microbiol Biotechnol.* 84(5):847-65.

*Rhodotorula* spp.

**[00140]** *Rhodotorula* is unicellular, pigmented yeast. The oleaginous red yeast, *Rhodotorula glutinis*, has been shown to produce lipids and carotenoids from crude glycerol (Saenge et al., 2011, *Process Biochemistry* 46(1):210-8). *Rhodotorula toruloides* strains have been shown to be an efficient fed-batch fermentation system for improved biomass and lipid productivity (Li et al., 2007, *Enzyme and Microbial Technology* 41:312-7).

*Rhodospiridium toruioides*

[00141] *Rhodospiridium toruioides* is oleaginous yeast and useful for engineering lipid-production pathways (See e.g. Zhu *et al.*, 2013, *Nature Commun.* 3:1 112; Ageitos *et al.*, 2011, *Applied Microbiology and Biotechnology* 90(4):1219-27).

*Candida boidinii*

[00142] *Candida boidinii* is methylotrophic yeast (it can grow on methanol). Like other methylotrophic species such as *Hansenula polymorpha* and *Pichia pastoris*, it provides an excellent platform for producing heterologous proteins. Yields in a multigram range of a secreted foreign protein have been reported. A computational method, IPRO, recently predicted mutations that experimentally switched the cofactor specificity of *Candida boidinii* xylose reductase from NADPH to NADH. See, e.g., Mattanovich *et al.*, 2012, *Methods Mol Biol.* 824:329-58; Khoury *et al.*, 2009, *Protein Sci.* 18(10):2125-38.

*Hansenula polymorpha* (*Pichia anomala*)

[00143] *Hansenula polymorpha* is methylotrophic yeast (see *Candida boidinii*). It can furthermore grow on a wide range of other substrates; it is thermo-tolerant and can assimilate nitrate (see also *Kluyveromyces lactis*). It has been applied to producing hepatitis B vaccines, insulin and interferon alpha-2a for the treatment of hepatitis C, furthermore to a range of technical enzymes. See, e.g., Xu *et al.*, 2014, *Virology* 509(2):403-9.

*Kluyveromyces lactis*

[00144] *Kluyveromyces lactis* is yeast regularly applied to the production of kefir. It can grow on several sugars, most importantly on lactose which is present in milk and whey. It has successfully been applied among others for producing chymosin (an enzyme that is usually present in the stomach of calves) for producing cheese. Production takes place in fermenters on a 40,000 L scale. See, e.g., van Ooyen *et al.*, 2006, *FEMS Yeast Res.* 6(3):381-92.

*Pichia pastoris*

[00145] *Pichia pastoris* is methylotrophic yeast (see *Candida boidinii* and *Hansenula polymorpha*). It provides an efficient platform for producing foreign proteins. Platform elements are available as a kit and it is worldwide used in academia for producing proteins. Strains have been engineered that can produce complex human N-glycan (yeast glycans are similar but not identical to those found in humans). See, e.g., Piirainen *et al.*, 2014, *N Biotechnol.* 31(6):532-7.

*Physcomitrella* spp.

**[00146]** *Physcomitrella* mosses, when grown in suspension culture, have characteristics similar to yeast or other fungal cultures. This genera can be used for producing plant secondary metabolites, which can be difficult to produce in other types of cells.

### **Steviol Glycoside Compositions**

**[00147]** Steviol glycosides do not necessarily have equivalent performance in different food systems. It is therefore desirable to have the ability to direct the synthesis to steviol glycoside compositions of choice. Recombinant hosts described herein can produce compositions that are selectively enriched for specific steviol glycosides (e.g., RebD or RebM) and have a consistent taste profile. As used herein, the term "enriched" is used to describe a steviol glycoside composition with an increased proportion of a particular steviol glycoside, compared to a steviol glycoside composition (extract) from a stevia plant. Thus, the recombinant hosts described herein can facilitate the production of compositions that are tailored to meet the sweetening profile desired for a given food product and that have a proportion of each steviol glycoside that is consistent from batch to batch. In some embodiments, hosts described herein do not produce or produce a reduced amount of undesired plant by-products found in *Stevia* extracts. Thus, steviol glycoside compositions produced by the recombinant hosts described herein are distinguishable from compositions derived from *Stevia* plants.

**[00148]** The amount of an individual steviol glycoside (e.g., RebA, RebB, RebD, or RebM) accumulated can be from about 1 to about 7,000 mg/L, e.g., about 1 to about 10 mg/L, about 3 to about 10 mg/L, about 5 to about 20 mg/L, about 10 to about 50 mg/L, about 10 to about 100 mg/L, about 25 to about 500 mg/L, about 100 to about 1,500 mg/L, or about 200 to about 1,000 mg/L, at least about 1,000 mg/L, at least about 1,200 mg/L, at least about at least 1,400 mg/L, at least about 1,600 mg/L, at least about 1,800 mg/L, at least about 2,800 mg/L, or at least about 7,000 mg/L. In some aspects, the amount of an individual steviol glycoside can exceed 7,000 mg/L. The amount of a combination of steviol glycosides (e.g., RebA, RebB, RebD, or RebM) accumulated can be from about 1 mg/L to about 7,000 mg/L, e.g., about 200 to about 1,500, at least about 2,000 mg/L, at least about 3,000 mg/L, at least about 4,000 mg/L, at least about 5,000 mg/L, at least about 6,000 mg/L, or at least about 7,000 mg/L. In some aspects, the amount of a combination of steviol glycosides can exceed 7,000 mg/L. In general, longer culture times will lead to greater amounts of product. Thus, the recombinant microorganism can be cultured for from 1 day to 7 days, from 1 day to 5 days, from 3 days to 5 days, about 3 days, about 4 days, or about 5 days.

[00149] It will be appreciated that the various genes and modules discussed herein can be present in two or more recombinant microorganisms rather than a single microorganism. When a plurality of recombinant microorganisms is used, they can be grown in a mixed culture to produce steviol and/or steviol glycosides. For example, a first microorganism can comprise one or more biosynthesis genes for producing a steviol glycoside precursor, while a second microorganism comprises steviol glycoside biosynthesis genes. The product produced by the second, or final microorganism is then recovered. It will also be appreciated that in some embodiments, a recombinant microorganism is grown using nutrient sources other than a culture medium and utilizing a system other than a fermenter.

[00150] Alternatively, the two or more microorganisms each can be grown in a separate culture medium and the product of the first culture medium, *e.g.*, steviol, can be introduced into second culture medium to be converted into a subsequent intermediate, or into an end product such as RebA. The product produced by the second, or final microorganism is then recovered. It will also be appreciated that in some embodiments, a recombinant microorganism is grown using nutrient sources other than a culture medium and utilizing a system other than a fermenter.

[00151] Steviol glycosides and compositions obtained by the methods disclosed herein can be used to make food products, dietary supplements and sweetener compositions. See, *e.g.*, WO 2011/153378, WO 2013/022989, WO 2014/122227, and WO 2014/122328.

[00152] For example, substantially pure steviol or steviol glycoside such as RebM or RebD can be included in food products such as ice cream, carbonated beverages, fruit juices, yogurts, baked goods, chewing gums, hard and soft candies, and sauces. Substantially pure steviol or steviol glycoside can also be included in non-food products such as pharmaceutical products, medicinal products, dietary supplements and nutritional supplements. Substantially pure steviol or steviol glycosides may also be included in animal feed products for both the agriculture industry and the companion animal industry. Alternatively, a mixture of steviol and/or steviol glycosides can be made by culturing recombinant microorganisms separately, each producing a specific steviol or steviol glycoside, recovering the steviol or steviol glycoside in substantially pure form from each microorganism and then combining the compounds to obtain a mixture comprising each compound in the desired proportion. The recombinant microorganisms described herein permit more precise and consistent mixtures to be obtained compared to current *Stevia* products.

**[00153]** In another alternative, a substantially pure steviol or steviol glycoside can be incorporated into a food product along with other sweeteners, e.g. saccharin, dextrose, sucrose, fructose, erythritol, aspartame, sucralose, monatin, or acesulfame potassium. The weight ratio of steviol or steviol glycoside relative to other sweeteners can be varied as desired to achieve a satisfactory taste in the final food product. See, e.g., U.S. 2007/0128311. In some embodiments, the steviol or steviol glycoside may be provided with a flavor (e.g., citrus) as a flavor modulator.

**[00154]** Compositions produced by a recombinant microorganism described herein can be incorporated into food products. For example, a steviol glycoside composition produced by a recombinant microorganism can be incorporated into a food product in an amount ranging from about 20 mg steviol glycoside/kg food product to about 1800 mg steviol glycoside/kg food product on a dry weight basis, depending on the type of steviol glycoside and food product. For example, a steviol glycoside composition produced by a recombinant microorganism can be incorporated into a dessert, cold confectionary (e.g., ice cream), dairy product (e.g., yogurt), or beverage (e.g., a carbonated beverage) such that the food product has a maximum of 500 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism can be incorporated into a baked good (e.g., a biscuit) such that the food product has a maximum of 300 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism can be incorporated into a sauce (e.g., chocolate syrup) or vegetable product (e.g., pickles) such that the food product has a maximum of 1000 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism can be incorporated into a bread such that the food product has a maximum of 160 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a hard or soft candy such that the food product has a maximum of 1600 mg steviol glycoside/kg food on a dry weight basis. A steviol glycoside composition produced by a recombinant microorganism, plant, or plant cell can be incorporated into a processed fruit product (e.g., fruit juices, fruit filling, jams, and jellies) such that the food product has a maximum of 1000 mg steviol glycoside/kg food on a dry weight basis. In some embodiments, a steviol glycoside composition produced herein is a component of a pharmaceutical composition. See, e.g., Steviol Glycosides Chemical and Technical Assessment 69th JECFA, 2007, prepared by Harriet Wailin, Food Agric. Org.; EFSA Panel on Food Additives and Nutrient Sources added to Food (ANS), "Scientific Opinion on the safety of steviol glycosides for the proposed uses as a food additive," 2010, *EFSA Journal* 8(4): 1537;

U.S. Food and Drug Administration GRAS Notice 323; U.S Food and Drug Administration GRAS Notice Notice 329; WO 201 1/037959; WO 2010/146463; WO 201 1/046423; and WO 2011/056834.

[001 55] For example, such a steviol glycoside composition can have from 90-99 weight % RebA and an undetectable amount of stevia plant-derived contaminants, and be incorporated into a food product at from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1 000 mg/kg on a dry weight basis.

[00156] Such a steviol glycoside composition can be a RebB-enriched composition having greater than 3 weight % RebB and be incorporated into the food product such that the amount of RebB in the product is from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1 000 mg/kg on a dry weight basis. Typically, the RebB-enriched composition has an undetectable amount of stevia plant-derived contaminants.

[001 57] Such a steviol glycoside composition can be a RebD-enriched composition having greater than 3 weight % RebD and be incorporated into the food product such that the amount of RebD in the product is from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1000 mg/kg on a dry weight basis. Typically, the RebD-enriched composition has an undetectable amount of stevia plant-derived contaminants.

[00158] Such a steviol glycoside composition can be a RebE-enriched composition having greater than 3 weight % RebE and be incorporated into the food product such that the amount of RebE in the product is from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1000 mg/kg on a dry weight basis. Typically, the RebE-enriched composition has an undetectable amount of stevia plant-derived contaminants.

[00159] Such a steviol glycoside composition can be a RebM-enriched composition having greater than 3 weight % RebM and be incorporated into the food product such that the amount of RebM in the product is from 25-1600 mg/kg, e.g., 100-500 mg/kg, 25-100 mg/kg, 250-1000 mg/kg, 50-500 mg/kg or 500-1000 mg/kg on a dry weight basis. Typically, the RebM-enriched composition has an undetectable amount of stevia plant-derived contaminants.

[00160] In some embodiments, a substantially pure steviol or steviol glycoside is incorporated into a tabletop sweetener or "cup-for-cup" product. Such products typically are diluted to the appropriate sweetness level with one or more bulking agents, e.g., maltodextrins, known to those skilled in the art. Steviol glycoside compositions enriched for RebA, RebB, RebD, RebE, or RebM, can be package in a sachet, for example, at from 10,000 to 30,000 mg

steviol glycoside/kg product on a dry weight basis, for tabletop use. In some embodiments, a steviol glycoside produced *in vitro*, *in vivo*, or by whole cell byconversion

[00161] The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

## EXAMPLES

[00162] The Examples that follow are illustrative of specific embodiments of the invention, and various uses thereof. They are set forth for explanatory purposes only, and are not to be taken as limiting the invention.

### Example 1. LC-MS Analytical Procedures

[00163] Three LC-MS procedures were used herein. In the first method used for Examples 2-6, LC-MS analyses were performed using an Ultimate 3000 UPLC system (Dionex) fitted with a Waters Acquity UPLC @BEH shield RP18 column (2.1 x 50 mm, 1.7  $\mu$ m particles, 130 Å pore size) connected to a TSQ Quantum Access (ThermoFisher Scientific) triple quadrupole mass spectrometer with a heated electrospray ion (HESI) source. Elution was carried out using a mobile phase of eluent B (MeCN with 0.1% formic acid) and eluent A (water with 0.1% formic acid) by increasing the gradient from 25% to 47% B from min 0.0 to 4.0, increasing 47% to 100% B from min 4.0 to 5.0, and holding 100% B from min 5.0 to 6.5. The flow rate was 0.4 mL/min and the column temperature 35°C. Steviol glycosides were detected using SIM (Single Ion Monitoring) with the following m/z-traces.

**Table 1A: LC-MS analytical information for Steviol Glycosides.**

Description	Exact Mass	m/z trace (Da)	compound (typical $t_R$ in min)
Steviol + 1 Glucose	[M+H] <sup>+</sup> 481.2796 [M+Na] <sup>+</sup> 503.2615	481.2 ± 0.5 503.1 ± 0.5	19-SMG (2.29), 13-SMG (3.5)
Steviol + 2 Glucose	[M+Na] <sup>+</sup> 665.3149	665 ± 0.5	Rubusoside (2.52) Steviol-1,2-bioside (2.92) Steviol-1,3-bioside (2.28)
Steviol + 3 Glucose	[M+Na] <sup>+</sup> 827.3677	827.4 ± 0.5	1,2-Stevioside (2.01) 1,3-Stevioside (2.39) Rebaudioside B (2.88)
Steviol + 4 Glucose	[M+Na] <sup>+</sup> 989.4200	989.4 ± 0.5	Rebaudioside A (2.0)
Steviol + 5 Glucose	[M+Na] <sup>+</sup> 1151.4728	1151.4 ± 0.5	Rebaudioside D (1.1)
Steviol +	[M+Na] <sup>+</sup> 1313.5257	1313.5 ± 0.5	Rebaudioside M (1.3)

Description	Exact Mass	m/z trace (Da)	compound (typical $t_R$ in min)
6 Glucose			

[00164] in the second method used for Examples 7, 8, and 10, LC-MS analyses were performed on Waters ACQUITY UPLC (Waters Corporation, Milford, MA) with coupled to a Waters ACQUITY ESI (electrospray ionization)-TQD triple quadrupole mass spectrometer. Compound separation was achieved on Waters ACQUITY UPLC® BEH C18 column (2.1 x 50 mm, 1.7  $\mu$ m particles, 130 Å pore size) equipped with ACQUITY UPLC BEH C18 VanGuard pre-column (130 Å, 1.7  $\mu$ m, 2.1 mm X 5 mm) by using a gradient of the two mobile phases: A (Water with 0.1% formic acid) and B (Acetonitrile with 0.1% formic acid) increasing B from 20% to 50% between 0.3 to 2.0 min up to 100% at 2.01 min, holding to 100% for 0.6 min, and re-equilibrating for 0.6 min. The flow rate was 0.6 mL/min, and the column temperature was 55°C. The MS acquisition was in negative ion-mode using SIM mode (Single Ion Monitoring). Steviol glycoside quantification was done by comparison with authentic standards.

Table 1B: MS analytical information for Steviol Glycosides.

Compound	m/z trace (Da)	Retention time (min)
RebE	965.42	1.06
RebD	1127.48	1.09
RebM	1289.53	1.15
RebA	965.42	1.43
1,3-Stevioside	803.37	1.60
Rubusoside	641.32	1.67
RebB	803.37	1.76
1,2-bioside	641.32	1.77
13-SMG	479.26	2.04

[00165] in the third method used for Example 9, LC-MS analyses were performed on Waters ACQUITY UPLC (Waters Corporation, Milford, MA) using a Waters Acquity UPLC® BEH C18 column (2.1 x 50 mm, 1.7  $\mu$ m particles, 130 Å) coupled to a Waters single quadrupole mass spectrometer (SQD), equipped with an ESI and operated in negative mode. Compound separation was achieved by a gradient of the two mobile phases: A (water with 0.1% formic acid) and B (acetonitrile with 0.1% formic acid) by increasing from 60% to 100% B between 0.3 to 2.5 min, holding 100% B for 0.1 min, and re-equilibrating for 0.2 min. The flow rate was 0.6 mL/min, and the column temperature was set at 55°C. Steviol or ent-kaurenoic acid was



monitored using SIM (Single Ion Monitoring) and quantified by comparing with authentic standards.

**Table 1C: MS analytical information for steviol and ent-kaurenoic acid.**

Compound	m/z trace (Da)	Retention time (min)
Steviol	317.21	0.61
Ent-kaurenoic acid	301.001	1.46

**Example 2. Construction of Steviol Glycoside-Producing and RebB-Producing Yeast Strains**

**[00166]** Steviol glycoside-producing *S. cerevisiae* strains were constructed as described in WO 201 1/153378, WO 2013/022989, WO 2014/122227, and WO 2014/122328. For example, a yeast strain comprising a recombinant gene encoding a *Synechococcus* sp. GGPPS (SEQ ID NO:49) polypeptide, a recombinant gene encoding a truncated *Zea mays* CDPS (SEQ ID NO:37) polypeptide, a recombinant gene encoding an *A. thaliana* KS (SEQ ID NO:6) polypeptide, a recombinant gene encoding an *S. rebaudiana* KO (SEQ ID NO:59, SEQ ID NO:79) polypeptide, a recombinant gene encoding an *A. thaliana* ATR2 (SEQ ID NO:51, SEQ ID NO:87) polypeptide, a recombinant gene encoding an *O. sativa* EUGT1 1 (SEQ ID NO:86) polypeptide, a recombinant gene encoding an SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) polypeptide, a recombinant gene encoding an *S. rebaudiana* CPR8 (SEQ ID NO:24, SEQ ID NO:28) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT85C2 (SEQ ID NO:30) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT74G1 (SEQ ID NO:29) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT76G1 (SEQ ID NO:2) polypeptide, and a recombinant gene encoding an *S. rebaudiana* UGT91D2 variant, UGT91D2e-b (SEQ ID NO:88), polypeptide accumulated steviol glycosides.

**[00167]** The UGT91D2e-b variant of UGT91D2 (SEQ ID NO:5 from PCT/US201 2/050021) includes a substitution of a methionine for leucine at position 211 and a substitution of an alanine for valine at position 286. Additional variants can include variants (except T144S, M152L, L213F, S364P, and G384C variants) described in Table 14 and Example 11 of the PCT/US201 2/050021. GeneArt codon-optimized sequence encoding a *S. rebaudiana* UGT91D2e-b with the amino acid modifications L211M and V286A (SEQ ID NO:88 for amino acid sequence; codon optimized nucleotide sequence is set forth in SEQ ID NO:89) and

expressed from the native yeast TDH3 promoter and followed by the native yeast CYC1 terminator.

**[00168]** Cells were grown in Synthetic Complete (SC) medium at 30°C for 5 days with shaking (400 rpm for deep wells and 200 rpm for 15 ml\_ Falcon growth tubes) prior to harvest. Culture samples (without cell removal) were heated in the presence of DMSO for detection of total glycoside levels with LC-MS. The strain accumulated total amounts of RebD of over 2500 mg/L, total amounts of RebM of over 2500 mg/L, and total amounts of RebA of over 700 mg/L. See WO 2014/122227.

**[00169]** A separate *S. cerevisiae* strain was constructed to accumulate RebB. This strain comprised a recombinant gene encoding a *Synechococcus* sp. GGPPS (SEQ ID NO:49) polypeptide, a recombinant gene encoding a truncated *Z. mays* CDPS (SEQ ID NO:37) polypeptide, a recombinant gene encoding an *A. thaliana* KS (SEQ ID NO:6) polypeptide, a recombinant gene encoding an *S. rebaudiana* KO (SEQ ID NO:59, SEQ ID NO:79) polypeptide, a recombinant gene encoding an *A. thaliana* ATR2 (SEQ ID NO:51, SEQ ID NO:87) polypeptide, a recombinant gene encoding an *O. sativa* EUGT1 1 (SEQ ID NO:86) polypeptide, a recombinant gene encoding an SrKAHei (SEQ ID NO:18, SEQ ID NO:68) polypeptide, a recombinant gene encoding an *S. rebaudiana* CPR8 (SEQ ID NO:24, SEQ ID NO:28) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT85C2 (SEQ ID NO:30) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT76G1 (SEQ ID NO:2) polypeptide, and a recombinant gene encoding an *S. rebaudiana* UGT91D2 variant, UGT91D2e-b (SEQ ID NO:88), polypeptide accumulated steviol glycosides.

### **Example 3. Steviol Glycoside Production in Yeast Strains Expressing KO Genes**

**[00170]** To determine whether increased levels of ent-kaurenoic acid improve steviol glycoside production, the activity of KO genes from various species were analyzed. Putative KO genes were identified using the NCBI Basic Local Alignment Sequence Search Tool (BLAST). Genes encoding KO polypeptides were cloned and expressed the RebB-producing *S. cerevisiae* strain described in Example 2, which was modified to lack KO genes. Thus, RebB was only accumulated upon expression of a functional KO.

**[00171]** Two KO polypeptides identified by the amino acid sequences set forth in SEQ ID NO:54 (nucleotide sequence set forth in SEQ ID NO:55) and SEQ ID NO:75 (nucleotide sequences set forth in SEQ ID NO:56) were found to accumulate higher levels of RebB than

SrKOl (nucleotide sequence set forth in SEQ ID NO:59, amino acid sequences set forth in SEQ ID NO:79) in the RebB-producing strain. RebB levels ( $\mu\text{M}/00_{600}$ ) are shown in Figure 3.

**[00172]** Expression of genes (SEQ ID NO:55 or SEQ ID NO:56) encoding KO polypeptides in an *S. cerevisiae* steviol glycoside-producing strain also resulted in accumulation of ent-kaurenoic acid (Figure 4). Expression of a gene encoding a codon-optimized KO polypeptide (SEQ ID NO:57) and a gene encoding the KO polypeptide set forth in SEQ ID NO:70 also resulted in accumulation of ent-kaurenoic acid. However, expression of SrKOl (SEQ ID NO:59, SEQ ID NO:79) did not result in measurable levels of ent-kaurenoic acid. Thus, the KO polypeptides encoded by nucleotide sequences set forth in SEQ ID NOs: 55-57 more efficiently converted ent-kaurene, ent-kaurenol, and/or ent-kaurenal to ent-kaurenoic acid in *S. cerevisiae*, as compared to the SrKOl polypeptide encoded by nucleotide sequence set forth in SEQ ID NO:59.

#### **Example 4. Steviol Glycoside Production in Yeast Strains Expressing KO Genes and Further Overexpressing SrKAHel**

**[00173]** Cloned KO genes were individually expressed in a steviol glycoside-producing *S. cerevisiae* strain. The *S. cerevisiae* strain described in Example 2, which expresses SrKOl (SEQ ID NO:59, SEQ ID NO:79), was modified to comprise overexpress SrKAHel (SEQ ID NO:18, SEQ ID NO:68). The coding sequences of the KO genes tested, as well as their corresponding amino acid sequences, are set forth in Table 2. The sequences set forth in SEQ ID NOs: 55, 57, 58, 59, and 60 were codon-optimized for expression in *S. cerevisiae*.

**Table 2: KO Genes Expressed in Steviol Glycoside-Producing *S. cerevisiae* strain that Further Overexpresses SrKAHel.**

KO Nucleotide Sequence	Corresponding KO Amino Acid Sequence
SEQ ID NO:55	SEQ ID NO:54
SEQ ID NO:56	SEQ ID NO:75
SEQ ID NO:57	SEQ ID NO:70
SEQ ID NO:58	SEQ ID NO:71
SEQ ID NO:59	SEQ ID NO:79
SEQ ID NO:60	SEQ ID NO:72

**[00174]** *S. cerevisiae* strains co-expressing any of the heterologous nucleic acids encoding a KO enzyme of Table 2 and further overexpressing SrKAHel (SEQ ID NO:18, SEQ ID NO:68)

accumulated higher levels of steviol glycosides than the control *S. cerevisiae* strain (not expressing a KO of Table 2) or a steviol glycoside-producing *S. cerevisiae* strain only overexpressing SrKAHel, as shown in Figure 5. A steviol glycoside-producing *S. cerevisiae* strain expressing a codon-optimized version of SEQ ID NO:56, identified herein as SEQ ID NO:65, and overexpressing SrKAHel accumulated higher levels of steviol glycosides (RebA, RebD, and RebM) than the steviol glycoside-producing *S. cerevisiae* strain co-expressing the nucleic acid set forth in SEQ ID NO:56 and SrKAHel (Figure 6).

**[00175]** Additionally, *S. cerevisiae* strains co-expressing a nucleic acid set forth in SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, or SEQ ID NO:60 and further overexpressing SrKAHel accumulated higher levels of glycosylated ent-kaurenoic acid than the control *S. cerevisiae* strain not expressing a KO of Table 2 (Figure 7).

**[00176]** As well, *S. cerevisiae* strains co-expressing a nucleic acid set forth in SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:59, or SEQ ID NO:60 and further overexpressing SrKAHel demonstrated improved metabolic conversion of intermediate compound, ent-kaurenol, which, in turn, resulted in reduced accumulation of glycosylated ent-kaurenol, relative to the control *S. cerevisiae* strain not expressing a KO of Table 2 or the steviol glycoside-producing *S. cerevisiae* strain only overexpressing SrKAHel, as shown in Figure 8. The control *S. cerevisiae* strain and the steviol glycoside-producing *S. cerevisiae* strain only overexpressing SrKAHel each accumulated higher levels of glycosylated ent-kaurenol than did *S. cerevisiae* strains expressing a nucleic acid set forth in SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:59, or SEQ ID NO:60 and further overexpressing SrKAHel.

#### **Example 5. Steviol Glycoside Production in Yeast Strains Expressing CPR Genes**

**[00177]** Cloned CPR genes were individually expressed in a steviol glycoside-producing *S. cerevisiae* strain. The steviol glycoside-producing *S. cerevisiae* strain described in Example 2, which expresses *S. rebaudiana* CPR8 (SEQ ID NO:24, SEQ ID NO:28) and *A. thaliana* ATR2 (SEQ ID NO:51), was modified to co-express a nucleic acid encoding a CPR of Table 3. The coding sequences of the CPR genes tested, as well as their corresponding amino acid sequences, are set forth in Table 3.

**Table 3: CPR Genes Tested in Combination with CPR8 and ATR2.**

Gene	Nucleotide Sequence	Amino Acid Sequence
------	---------------------	---------------------

<i>S. rebaudiana</i> CPR1	SEQ ID NO:61	SEQ ID NO:76
<i>S. rebaudiana</i> CPR7	SEQ ID NO:23	SEQ ID NO:69
CPR4497	SEQ ID NO:62	SEQ ID NO:74

**[00178]** As shown in Figure 9, expression of CPR1 (SEQ ID NO:61, SEQ ID NO:76) or of CPR7 (SEQ ID NO:23, SEQ ID NO:69) in the steviol glycoside-producing *S. cerevisiae* strain already expressing *S. rebaudiana* CPR8 (SEQ ID NO:24, SEQ ID NO:28) and *A. thaliana* ATR2 (SEQ ID NO:51) resulted in higher levels of RebM than those accumulated by the control steviol glycoside-producing *S. cerevisiae* strain not expressing CPR1 or CPR7. As well, a steviol glycoside-producing *S. cerevisiae* strain expressing the nucleic acid set forth in SEQ ID NO:62 and overexpressing SrKAHel (SEQ ID NO:18, SEQ ID NO:68) accumulated higher levels of RebM than those accumulated by the control steviol glycoside-producing *S. cerevisiae* strain that only overexpressed SrKAHel (Figure 10).

#### **Example 6. Steviol Glycoside Production in Yeast Strains Co-Expressing KO and CPR Genes**

**[00179]** Steviol glycoside production was tested in the RebB-producing *S. cerevisiae* strain described in Example 2, which was modified to co-express a KO gene of Table 4 and a CPR of Table 5.

**Table 4: KO Genes Tested in Combination with CPR Genes.**

Gene	Nucleotide Sequence	Amino Acid Sequence
SrKO1	SEQ ID NO:59	SEQ ID NO:79
Codon-optimized KO	SEQ ID NO:63	SEQ ID NO:77
Codon-optimized KO	SEQ ID NO:64	SEQ ID NO:78

**Table 5: CPR Genes Tested in Combination with KO Genes.**

Nucleotide Sequence	Amino Acid Sequence
SEQ ID NO:66	SEQ ID NO:73
SEQ ID NO:67	SEQ ID NO:22

**[00180]** As shown in Figure 12, co-expression of SrKO1 (SEQ ID NO:59, SEQ ID NO:79) and either of the CPR genes of Table 5 in the RebB-producing strain resulted in higher production of 13-SMG and RebB than co-expression of a nucleic acid set forth in SEQ ID NO:63 or SEQ ID NO:64 and either of the cytochrome P450 genes of Table 5.

**Example 7. Steviol Glycoside Production in Yeast Strains Expressing KAH Genes**

**[00181]** Candidate KAH enzymes were cloned and expressed in an *S. cerevisiae* strain engineered to accumulate 13-SMG. The 13-SMG-producing *S. cerevisiae* strain comprised a recombinant gene encoding a *Synechococcus* sp. GGPPS7 polypeptide (SEQ ID NO:49), a recombinant gene encoding a truncated *Z. mays* CDPS polypeptide (SEQ ID NO:37), a recombinant gene encoding an *A. thaliana* KS polypeptide (SEQ ID NO:6), SrKOI (SEQ ID NO:59, SEQ ID NO:79), CPR8 (SEQ ID NO:24, SEQ ID NO:28), the KO encoded by the nucleotide sequence set forth in SEQ ID NO:56 (amino acid sequence set forth in SEQ ID NO:75), and UGT85C2 (SEQ ID NO:30) chromosomally integrated in separate expression cassettes (Figure 11B). The strain lacked SrKAHeI (SEQ ID NO:18, SEQ ID NO:68); thus, 13-SMG was only accumulated upon transformation of the *S. cerevisiae* strain with a functional KAH (Figure 11B).

**[00182]** Transformants were grown in SC-URA medium for 4 days and extracted with 1:1 with DMSO at 80°C for 10 min. The extracts were analyzed by LC-MS (method 2 of Example 1). *S. cerevisiae* transformed with the nucleic acid set forth in SEQ ID NO:80 accumulated 13-SMG (Figure 11B). Thus, the protein encoded by SEQ ID NO:80, set forth in SEQ ID NO:82, is a KAH.

**[00183]** The KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80 was codon-optimized for expression in yeast (SEQ ID NO:81) and expressed in the above-described 13-SMG-producing *S. cerevisiae* strain. Similar to expression of SrKAHeI (SEQ ID NO:18) or the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80, expression of the codon-optimized nucleotide sequence set forth in SEQ ID NO:81 resulted in production of 13-SMG plus rubusoside (Figure 13).

**[00184]** The KAHs encoded by the nucleotide sequence set forth in SEQ ID NO:80 and the codon-optimized nucleotide sequence set forth in SEQ ID NO:81 were also individually expressed in a steviol glycoside-producing strain, as described in Example 2, which expresses SrKAHeI. Production of 13-SMG was increased upon overexpression of SrKAHeI (SEQ ID NO:18), of the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80, or of the KAH encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:81, as compared to a control strain not expressing the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80, the KAH encoded by the codon-optimized nucleotide sequence set forth

in SEQ ID NO:81, or overexpressing SrKAHe1. See Table 6. Expression of either the KAH encoded by the nucleotide sequence set forth in SEQ ID NO:80 or the KAH encoded by the codon-optimized nucleotide sequence set forth in SEQ ID NO:81 resulted in higher steviol glycoside production (13-SMG + 1,2-bioside + rubusoside + RebB + RebA + RebD + RebM) than either the control strain or the *S. cerevisiae* strain overexpressing SrKAHe1 (SEQ ID NO: 18). See Table 6.

**Table 6: Quantification of Steviol Glycosides Accumulated by Yeast Expressing KAH Genes.**

	Control ( $\mu$ M)	Overexpression of SrKAHe1 (encoded by the nucleotide set forth in SEQ ID NO:18) ( $\mu$ M)	SrKAHe1 + KAH (encoded by the nucleotide set forth in SEQ ID NO:80) ( $\mu$ M)	SrKAHe1 + KAH (encoded by the nucleotide sequence set forth in SEQ ID NO:81) ( $\mu$ M)
<b>13-SMG</b>	67.6	85.5	153.8	130.5
<b>Steviol-1,2-bioside</b>	0.4	0.3	0.4	0.4
<b>Rubusoside</b>	1.2	1.0	1.4	1.1
<b>RebB</b>	8.6	7.6	9.6	9.6
<b>RebA</b>	30.7	26.0	26.8	28.7
<b>RebD</b>	36.2	27.6	32.9	36.5
<b>RebM</b>	138.3	118.9	100.0	90.3
<b>Sum</b>	282.7	266.2	324.0	296.7

**Example 8. Steviol Glycoside Production in Yeast Strain Expressing KAH Gene of the CYP72A219 family**

**[001 85]** A nucleic acid of SEQ ID NO:90, which was codon-optimized for expression in *S. cerevisiae* and encodes the polypeptide of SEQ ID NO:91, was cloned and expressed in an *S. cerevisiae* strain described in Example 7, which was engineered to accumulate 13-SMG. The 13-SMG-producing *S. cerevisiae* strain comprised a recombinant gene encoding a *Synechococcus* sp. GGPPS7 polypeptide (SEQ ID NO:49), a recombinant gene encoding a truncated *Z. mays* CDPS polypeptide (SEQ ID NO:37), a recombinant gene encoding an *A. thaliana* KS polypeptide (SEQ ID NO:6), SrKOI (SEQ ID NO:59, SEQ ID NO:79), CPR8 (SEQ ID NO:24, SEQ ID NO:28), the KO encoded by the nucleotide sequence set forth in SEQ ID NO:56 (amino acid sequence set forth in SEQ ID NO:75), and UGT85C2 (SEQ ID NO:30) chromosomally integrated in separate expression cassettes.

**[00186]** Transformants were grown in SC-URA medium for 4 days and extracted 1:1 with DMSO at 80°C for 10 min. The extracts were analyzed by LC-MS (method 2 of Example 1). *S. cerevisiae* transformed with the nucleic acid set forth in SEQ ID NO:90 accumulated 13-SMG as well as rubusoside (Table 7). Thus, the protein encoded by the nucleic acid sequence of SEQ ID NO:90, set forth in SEQ ID NO:91, is a KAH.

**Table 7: Quantification of Steviol Glycosides Accumulated by Yeast Expressing the KAH encoded by the Nucleotide Sequence Set Forth in SEQ ID NO:90 (Amino Acid Sequence Set Forth in SEQ ID NO:91).**

	13-SMG (μM)	Rubusoside (μM)
KAH (encoded by the nucleotide sequence set forth in SEQ ID NO:90)	4.3 ± 0.1	0.2 ± 0.0

#### **Example 9. Determination of CPR1 and CPR12 Activity**

**[00187]** Activity of CPR1 and CPR12 were measured using an *in vitro* microsomal assay. Microsomes were prepared by a modified version of the method taught by Pompon *et al.*, "Yeast expression of animal and plant P450s in optimized redox environments," *Methods Enzymol.* 272:51-64 (1996). *S. cerevisiae* cells were sedimented for 10 min at 4°C. The pellets were washed with 10 mL TEK buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM KCl.) The cells were sedimented again for 10 min at 4°C, and the pellets were resuspended in 1-3 mL of TES2 buffer (50 mM Tris-HCl (pH 7.5) 1 mM EDTA, 600 mM sorbitol). Glass beads (425-600 microns) were added to the samples, and the cells were broken vigorously by shaking and vortexing for 5 min at 4°C. The supernatant was collected, and the beads were washed several times with TES2 buffer. The washes were combined with the supernatant, and the samples were centrifuged for 15 min at 4°C to remove unbroken cells and glass beads. Samples were then ultracentrifuged for 1 h at 4°C. The pellets were washed twice with TES buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 600 mM sorbitol, 1% (w/v) BSA, 5 mM DTT), and once with TEG buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 30% (V/V) glycerol). The samples were resuspended in 1-3 mL TEG, and the pellets were homogenized.

**[00188]** Wild-type control microsomal protein was prepared as described above from wild-type *S. cerevisiae* cells that did not comprise a heterologous KAH or CPR. Microsomal protein



was also prepared from *S. cerevisiae* cells expressing i) SrKAHel (SEQ ID NO:18, SEQ ID NO:68), ii) SrKAHel (SEQ ID NO:18, SEQ ID NO:68) and CPR1 (SEQ ID NO:61, SEQ ID NO:76), or iii) SrKAHel (SEQ ID NO:18, SEQ ID NO:68) and CPR12 (SEQ ID NO:97, SEQ ID NO:98) from a genetic construct integrated at the chromosome level. Microsomal protein from a steviol glycoside-producing strain was prepared from *S. cerevisiae* cells expressing the genes described in Example 2 and additionally comprising codon-optimized CPR1 from *S. rebaudiana* (SEQ ID NO:61 corresponding to amino acid sequence SEQ ID NO:76) as well as the KO encoded by SEQ ID NO:75).

**[00189]** CPR1 and CPR12 activities were first determined using a cytochrome C reductase assay kit (Sigma-Aldrich; CY0100-1KT) to measure the ability of CPR1 or CPR12 to reduce cytochrome C in the presence of NADPH *in vitro*. Reduction of cytochrome C resulted in an increase in absorbance at 550 nm, which could be quantified spectrophotometrically. Working solution was prepared by adding 9 mg cytochrome C to 20 mL assay buffer, and solution was stored at 25°C until use. NADPH was diluted in H<sub>2</sub>O to a concentration of 0.85 mg/mL. Final reaction volumes were 1.1 mL (950 µL working solution (0.43 mg cytochrome C), 28 µL enzyme dilution buffer, 100 µL NADPH solution (0.085 mg NADPH), 20 µL cytochrome C oxidase inhibitor, 2 µL microsomal protein.) Blank samples did not comprise microsomal protein and were prepared with 950 µL working solution (0.43 mg cytochrome C), 30 µL enzyme dilution buffer, 100 µL NADPH solution (0.085 mg NADPH), and 20 µL cytochrome C oxidase inhibitor. The spectrophotometer was blanked with all components added to the reactions except for NADPH. The enzymatic reactions were initiated by addition of NADPH, the samples were thoroughly mixed by pipetting, and absorbance was measured at 550 nm for 70 s with 10 s intervals between reads. Two independent rate measurements were taken for each microsomal preparation, and rates were averaged for calculation of specific activity. After the reactions were completed, results were normalized to protein concentration, which was measured using a standard BCA assay (Thermo Scientific).

**[00190]** Units/mL was calculated using the following equation, where  $\Delta A_{550}/\text{min}$  represents the change in absorbance at 550 nm during the absorbance reading period, 1.1 represents the reaction volume in mL, and 21.1 represents the extinction coefficient for reduced cytochrome c:

$$\text{Units/mL} = (\Delta A_{550}/\text{min} \times \text{dilution factor} \times 1.1) / (21.1 \times \text{enzyme volume})$$

**[00191]** The units/mL value of each sample was divided by its respective microsomal protein concentrations to calculate CPR activity in units/mg. Figure 14 shows the activity measurements of the i) SrKAHel (SEQ ID NO:18, SEQ ID NO:68), ii) SrKAHel (SEQ ID NO:18,

SEQ ID NO:68) and CPR1 (SEQ ID NO:61, SEQ ID NO:76), and iii) SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) and CPR12 (SEQ ID NO:97, SEQ ID NO:98) microsomal samples.

**[00192]** The microsomal preparation from the wild-type control showed only minimal CPR activity, reflecting the low activity of native NCP1 (YHR042W). Likewise, the microsomal preparation from a yeast strain overexpressing KAHeI did not demonstrate an increase in CPR activity. In contrast, microsomal preparation from strains expressing SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) and CPR1 (SEQ ID NO:61, SEQ ID NO:76) or SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) and CPR12 (SEQ ID NO:97, SEQ ID NO:98) demonstrated high CPR activity, with 7- and 14-fold higher activity, respectively, compared to the negative control (Figure 14).

**[00193]** In a separate experiment, formation of steviol and consumption of ent-kaurenoic acid in microsomes, as prepared above, were measured. 33  $\mu$ M ent-kaurenoic acid, 10 mM NADPH, and 10  $\mu$ L of microsomal protein in 50 mM phosphate buffer (pH 7.5) were incubated for 30 min at 30°C in a total reaction volume of 100  $\mu$ L. Control reactions were extracted immediately after addition of all the reaction components, which were mixed on ice and aliquoted prior to incubation. Steviol and ent-kaurenoic acid levels were quantified using the second LC-MS procedure described in Example 1. For steviol quantification, the microsomal reactions were extracted with DMSO (1:1) at 80°C for 10 min and submitted for LC-MS analysis after centrifugation. For ent-kaurenoic acid quantification the microsomes reactions were extracted with acetonitrile 1:4 (20% microsomal reaction and 80% acetonitrile) at 80°C for 10 min and after centrifugation submitted for LC-MS analysis. The AUC values obtained for the ent-kaurenoic acid measurements were converted to concentrations using a standard curve.

**[00194]** As shown in Figure 15A, microsomal protein prepared from an *S. cerevisiae* strain expressing SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) and either CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) converted ent-kaurenoic acid to steviol during the 30 minute incubation period. The steviol level shown in Figure 15A for the steviol-glycoside-producing strain control (extracted immediately with no 30 min incubation period) corresponds to steviol that was accumulated by the strain prior to microsomal preparation and that had co-purified with the microsomes. As shown in Figure 15B, ent-kaurenoic acid levels decreased upon incubation with microsomal protein prepared from *S. cerevisiae* strains expressing SrKAHeI (SEQ ID NO:18, SEQ ID NO:68) alone or in combination with CPR1 (SEQ ID NO:61, SEQ ID NO:76) or CPR12 (SEQ ID NO:97, SEQ ID NO:98). The increased ent-kaurenoic acid levels shown in Figure 15B for the steviol glycoside-producing strain microsomal sample incubated for 30 min corresponds to ent-kaurenoic acid that was accumulated by the

strain prior to microsomal preparation and to ent-kaurenoic acid accumulated from ent-kaurene that had co-purified with the microsomes. The levels of ent-kaurenoic acid shown in Figure 15B were corrected for the dilution factor used.

**Example 10. Steviol Glycoside Production in *S. cerevisiae* strains comprising Fusion Constructs between a KO and a P450 Reductase Domain**

**[00195]** CYP102A1 (also referred to as P450<sub>BM3</sub>; SEQ ID NO:115, SEQ ID NO:116) is a catalytically self-sufficient soluble enzyme from *Bacillus megatarium*. See, e.g., Whitehouse *et al*, 2012, Chem Soc Rev. 41(3):1218-60. Two domains are present in the CYP102A1 polypeptide chain: a P450 heme domain (BMP) and an NADPH-dependent P450 oxidoreductase domain (BMR). CYP102A1 utilizes nearly 100% of the reducing power of NADPH to produce a monooxygenated product. See, e.g., Yuan *et al*, 2009, *Biochemistry* 48(38):9140-6.

**[00196]** The BMR domain of CYP102A1 ("BMR"; codon-optimized nucleotide sequence set forth in SEQ ID NO:117, SEQ ID NO:118) was fused to SrKOi (SEQ ID NO:59, SEQ ID NO:79) or a KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 (amino acid sequence set forth in SEQ ID NO:75) with a linker (SEQ ID NO:121, SEQ ID NO:122), as described in Dodhia *et al*, 2006, J Biol Inorg Chem. 11(7):903-16. A wild-type version of the BMR domain of CYP102A1, as well as a W1046A mutant of the BMR domain (SEQ ID NO:119, SEQ ID NO:120), which has been found to switch the cofactor specificity of CYP102A1 from NADPH to NADH, were used. See, Girvan *et al*., 2011, Arch Biochem Biophys. 507(1):75-85. SrKOi (SEQ ID NO:59, SEQ ID NO:79) and the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 were also truncated prior to fusion with the BMR domain of CYP102A1; these truncations were predicted by bioinformatics to result in loss of membrane anchors of the KO genes and in cytosolic versions of the KO-BMR fusion constructs. The KO-BMR fusion constructs analyzed are shown in Table 8.

**Table 8: KO-BMR fusion constructs and sequences.**

Fusion Construct	Codon-Optimized Nucleotide Sequence	Amino Acid Sequence
SrKO1-BMR	SEQ ID NO:99	SEQ ID NO:100
SrKO1-BMR W1046A mutant	SEQ ID NO:101	SEQ ID NO:102
Truncated SrKO1-BMR	SEQ ID NO:103	SEQ ID NO:104

Truncated SrKO1-BMR W1046A mutant	SEQ ID NO:105	SEQ ID NO:106
KO ( <i>encoded by nucleotide sequence set forth in SEQ ID NO:65</i> )-BMR	SEQ ID NO:107	SEQ ID NO:108
KO ( <i>encoded by nucleotide sequence set forth in SEQ ID NO:65</i> )-BMR W1046A mutant	SEQ ID NO:109	SEQ ID NO:110
Truncated KO ( <i>encoded by nucleotide sequence set forth in SEQ ID NO:65</i> )-BMR W1046A mutant	SEQ ID NO:111	SEQ ID NO:112

[00197] The KO-BMR fusion constructs were cloned and transformed in the RebB-producing strain described in Example 2, which was modified to not comprise any additional KO genes. Thus, steviol glycosides, including 13-SMG, 1,2-bioside, and RebB, were only accumulated upon expression of a functional KO. Three scrapes (1  $\mu$ L loop of cells) from each transformation plate were resuspended in 200  $\mu$ L nanopure H<sub>2</sub>O. 70  $\mu$ L were then transferred to 1 mL SC-URA in a 96 deep well plate and incubated at 30°C for 5 days at 400 rpm. Biological triplicates were analyzed by LC-MS (method 2 of Example 1) to measure 13-SMG, 1,2-bioside, and RebB levels, and single samples were analyzed by LC-UV to measure ent-kaurene and ent-kaurenoic acid levels.

[00198] For LC-MS, 50  $\mu$ L samples were mixed with 50  $\mu$ L 100% DMSO and heated to 80°C for 10 min. Subsequently, the samples were spun down at 4000 RCF for 10 min, and 85  $\mu$ L of the resulting supernatant was transferred to an LC-MS plate. The LC-MS results were normalized by OD<sub>600</sub> of individual cultures, which was measured by a Wallac, 2104 EnVision (Perkin Elmer) plate reader.

[00199] LC-UV was conducted with an Agilent 1290 instrument comprising a variable wavelength detector (VWD), a thermostatted column compartment (TCC), an autosampler, an autosampler cooling unit, and a binary pump and using SB-C18 rapid resolution high definition (RRHD) 2.1 mm x 300 mm, 1.8  $\mu$ m analytical columns (two 150 mm columns in series; column temperature of 65°C). Steviol glycosides and steviol glycoside precursors were separated by a reversed phase C18 column followed by detection by UV absorbance at 210 nm. Quantification of steviol glycosides was done by comparing the peak area of each analyte to standards of RebA and applying a correction factor for species with differing molar

absorptivities. Quantification of steviol glycoside precursors (such as kaurenoic acid, kaurenal, kaurenol, ent-kaurene, and geranylgeraniol) was done by comparing the peak area of each analyte to standards of kaurenoic acid and applying a correction factor for species with differing molar absorptivities. For LC-UV, 0.5 mL cultures were spun down, the supernatant was removed, and the wet weight of the pellets was calculated. The LC-UV results were normalized by pellet wet weight.

**[00200]** As shown in Figures 16B and 16D, the *S. cerevisiae* strain transformed with empty plasmid accumulated ent-kaurene. Transformation with a plasmid comprising SrKOI (SEQ ID NO:59, SEQ ID NO:79) or with a plasmid comprising the KO gene having the nucleotide sequence set forth in SEQ ID NO:65 resulted in accumulation of 13-SMG, 1,2-bioside, and RebB (Figures 16A and 186C).

**[00201]** Expression of full-length SrKOI -BMR fusion constructs (wild type or W1046A mutant BMR; SEQ ID NOs:99-102), resulted in an increase in ent-kaurenoic acid, 13-SMG, and RebB, compared to expression of SrKOI (SEQ ID NO:59, SEQ ID NO:79). See Figures 16A and 16B. Expression of truncated SrKOI -BMR fusion constructs (wild type or W1046A mutant BMR; SEQ ID NOs:103-106) resulted in an increase in ent-kaurenoic acid, compared to expression of SrKOI (SEQ ID NO:59, SEQ ID NO:79) (Figure 16B). Although the truncated SrKOI -BMR fusion constructs also increased steviol glycoside production, glycosylation activity was higher for the full-length SrKOI-BMR fusion constructs than for the truncated SrKOI -BMR fusion constructs (Figure 16A).

**[00202]** Expression of a fusion construct comprising the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and the wild type BMR (SEQ ID NO:107, SEQ ID NO:108) resulted in greater conversion of ent-kaurenoic acid to 13-SMG, compared to the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 (Figure 16C). Expression of a fusion construct comprising the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 and the W1046A mutant BMR (SEQ ID NO:109, SEQ ID NO:110) resulted in decreases in ent-kaurenoic acid levels but glycosylation activity similar to that of the KO encoded by the nucleotide sequence set forth in SEQ ID NO:65 (Figure 16C).

#### **Example 11. Evaluation of Steviol Glycoside Pathway in *S. cerevisiae* Strain Comprising ICE2**

**[00203]** ICE2 is an endoplasmic reticulum (ER) membrane protein involved in mechanisms such as ER zinc homeostasis and cytochrome P450 stability and/or activity. See, e.g., Estrada de Martin *et al.*, 2005, J Cell Sci. 118(Pt 1):65-77 and Emmerstorfer *et al.*, 2015, Biotechnol J. 10(4):623-35. ICE2 (SEQ ID NO:113, SEQ ID NO:114) was cloned and overexpressed in a steviol glycoside-producing *S. cerevisiae* strain comprising a recombinant gene encoding a *Synechococcus* sp. GGPPS polypeptide (SEQ ID NO:49), a recombinant gene encoding a truncated *Z. mays* CDPS polypeptide (SEQ ID NO:37), a recombinant gene encoding an *A. thaliana* KS polypeptide (SEQ ID NO:6), a recombinant gene encoding a recombinant *S. rebaudiana* KO polypeptide (SEQ ID NO:59, SEQ ID NO:79), a recombinant gene encoding an *A. thaliana* ATR2 polypeptide (SEQ ID NO:51, SEQ ID NO:87), a recombinant gene encoding an SrKAHel (SEQ ID NO:18, SEQ ID NO:68) polypeptide, a recombinant gene encoding an *S. rebaudiana* CPR8 polypeptide (SEQ ID NO:24, SEQ ID NO:28), a recombinant KAH gene encoded by the nucleotide sequence set forth in SEQ ID NO:81 (corresponding to the amino acid sequence set forth in SEQ ID NO:82), a recombinant KO gene encoded by the nucleotide sequence set forth in SEQ ID NO:56 (corresponding to the amino acid sequence set forth in SEQ ID NO:75), a recombinant KO gene encoded by the nucleotide sequence set forth in SEQ ID NO:65 (corresponding to the amino acid sequence set forth in SEQ ID NO:75), a recombinant gene encoding a UGT76G1 (SEQ ID NO:83) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT85C2 polypeptide (SEQ ID NO:30), a recombinant gene encoding an *S. rebaudiana* UGT74G1 polypeptide (SEQ ID NO:29), a recombinant gene encoding an EUGT11 (SEQ ID NO:86) polypeptide, a recombinant gene encoding a UGT91 D2e (SEQ ID NO:84) polypeptide, and a recombinant gene encoding a CPR1 (SEQ ID NO:61, SEQ ID NO:76) polypeptide. Overexpression was performed by integration using the USER cloning system; see, e.g., Nour-Eldin *et al.*, 2010, *Methods Mol Biol.* 643:185-200. Table 9 shows additional recombinant genes (ICE2 and/or CPR12) expressed in the above-described strain. The control strain did not comprise recombinant genes encoding ICE2 (SEQ ID NO:113, SEQ ID NO:114) or CPR12 (SEQ ID NO:97, SEQ ID NO:98) polypeptides.

**Table 9: ICE2 steviol glycoside-producing strains.**

Strain	Sequences
ICE2 "strain A"	ICE2 (SEQ ID NO:113, SEQ ID NO:114) Overexpressed CPR1 (SEQ ID NO:61, SEQ ID NO:76)
ICE2 "strain B"	ICE2 (SEQ ID NO:113, SEQ ID NO:114) (2 copies)

ICE2 "strain C"	ICE2 (SEQ ID NO:113, SEQ ID NO:114) CPR12 (SEQ ID NO:97, SEQ ID NO:98)
-----------------	---

**[00204]** Fed-batch fermentation was carried out aerobically in 2 L fermenters at 30°C with an approximate 16 h growth phase in minimal medium comprising glucose, ammonium sulfate, trace metals, vitamins, salts, and buffer followed by an approximate 110 h feeding phase with a glucose-comprising defined feed medium. A pH near 6.0 and glucose-limiting conditions were maintained. Whole culture samples (without cell removal) were analysed by the LC-UV method of Example 10 to determine levels of steviol glycosides and steviol pathway intermediates.

**[00205]** The following values were calculated based upon the measured levels of steviol glycosides and steviol glycoside precursors. "Total Flux" was calculated as a sum (in g/L RebD equivalents) of measured RebA, RebB, RebD, RebE, RebM, 13-SMG, rubusoside, steviol-1,2-bioside, di-glycosylated steviol, tri-glycosylated steviol, tetra-glycosylated steviol, penta-glycosylated steviol, hexa-glycosylated steviol, hepta-glycosylated steviol, copalol, ent-kaurenoic acid, glycosylated ent-kaurenoic acid, glycosylated ent-kaurenol, ent-kaurenol, geranylgeraniol, ent-kaurenol, and ent-kaurene levels. "Pre-steviol glycoside/flux" was calculated as  $((\text{"total flux"} - (\text{geranylgeraniol} + \text{copalol} + \text{ent-kaurene} + \text{glycosylated ent-kaurenol} + \text{ent-kaurenol} + \text{ent-kaurenol} + \text{ent-kaurenoic acid} + \text{glycosylated ent-kaurenoic acid})) / \text{"total flux"})$ . "KAH step/flux" was calculated as  $((\text{ent-kaurenoic acid} + \text{glycosylated ent-kaurenoic acid}) / \text{"total flux"})$ . "KO step/flux" was calculated as  $((\text{ent-kaurene} + \text{glycosylated ent-kaurenol} + \text{ent-kaurenol} + \text{ent-kaurenol}) / \text{"total flux"})$ .

**[00206]** The pre-steviol glycoside/flux, KO step/flux, and KAH step/flux values are shown in Table 10 below. Decreased amounts of ent-kaurene, ent-kaurenol, ent-kaurenol, glycosylated ent-kaurenol and increased amounts of ent-kaurenoic acid and glycosylated ent-kaurenoic acid were observed in the strains comprising ICE2, as compared to the control steviol glycoside-producing strain. These effects were stronger in the presence of CPR1 and/or CPR12 (Table 10). Overexpression of two copies of ICE2 (ICE2 strain B) resulted decreased ent-kaurene, ent-kaurenol, ent-kaurenol, and ent-kaurenol glycoside levels and increased steviol glycoside levels, compared to the control strain, ICE2 strain A, or ICE2 strain C (Table 10). Steviol glycoside levels increased most in the steviol glycoside-producing strain comprising two copies of ICE2. Thus, ICE2 was found to improve cytochrome P450 function.

**Table 10: Pre-steviol glycoside/flux, KO step/flux, and KAH step/flux values for steviol glycoside-producing strains comprising ICE2.**

Strain	Pre-Steviol Glycoside/Flux	KO step/Flux	KAH step/Flux
ICE2 "strain A"	0.38	0.36	0.22
ICE2 "strain B"	0.43	0.42	0.10
ICE2 "strain C"	0.39	0.38	0.19
Control	0.41	0.48	0.08

**Example 12. Steviol Glycoside Production by Fermentation of *S. cerevisiae* strain comprising CPR1 and CPR12**

**[00207]** Steviol glycoside-producing *S. cerevisiae* strains comprising a recombinant gene encoding a *Synechococcus* sp. GGPPS polypeptide (SEQ ID NO:49), a recombinant gene encoding a truncated *Z. mays* CDPS polypeptide (SEQ ID NO:37), a recombinant gene encoding an *A. thaliana* KS polypeptide (SEQ ID NO:6), a recombinant gene encoding a recombinant *S. rebaudiana* KO polypeptide (SEQ ID NO:59, SEQ ID NO:79), a recombinant gene encoding an *A. thaliana* ATR2 polypeptide (SEQ ID NO:51, SEQ ID NO:87), a recombinant gene encoding an SrKAHei (SEQ ID NO:18, SEQ ID NO:68) polypeptide, a recombinant gene encoding an *S. rebaudiana* CPR8 polypeptide (SEQ ID NO:24, SEQ ID NO:28), a recombinant gene encoding a CPR1 (SEQ ID NO:61, SEQ ID NO:76) polypeptide, a recombinant gene encoding an SrKAHei (SEQ ID NO:18, SEQ ID NO:68) polypeptide, a recombinant KO gene encoded by the nucleotide sequence set forth in SEQ ID NO:56 (corresponding to the amino acid sequence set forth in SEQ ID NO:75), a recombinant gene encoding a UGT76G1 (SEQ ID NO:83) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT85C2 (SEQ ID NO:30) polypeptide, a recombinant gene encoding an *S. rebaudiana* UGT74G1 (SEQ ID NO:29) polypeptide, a recombinant gene encoding a UGT91D2e-b polypeptide (SEQ ID NO:88), and a recombinant gene encoding an EUGT1 1 (SEQ ID NO:86) polypeptide, as well as the recombinant genes shown in Table 11, which were genomically integrated into the strains, were cultivated by fermentation. Levels of steviol glycosides and steviol glycoside precursors were measured by LC-UV as described in Example 11. The pre-KO/flux, pre-KAH/flux, pre-steviol glycoside/flux values were calculated as described in Example 11.

**Table 11: Recombinant genes also expressed in steviol glycoside-producing *S. cerevisiae* strain in Example 12.**

Strain	Genes
--------	-------



Example 12, Strain A	KO encoded by nucleotide sequence set forth in SEQ ID NO:56 (corresponding to amino acid sequence set forth in SEQ ID NO:75)
Example 12, Strain B	KAH encoded by nucleotide sequence set forth in SEQ ID NO:80 (corresponding to amino acid sequence set forth in SEQ ID NO:82)  KO encoded by nucleotide sequence set forth in SEQ ID NO:56 (corresponding to amino acid sequence set forth in SEQ ID NO:75)  KO encoded by nucleotide sequence set forth in SEQ ID NO:65 (corresponding to amino acid sequence set forth in SEQ ID NO:75)
Example 12, Strain C	CPR12 (SEQ ID NO:97, SEQ ID NO:98)  KAH encoded by nucleotide sequence set forth in SEQ ID NO:80 (corresponding to amino acid sequence set forth in SEQ ID NO:82)  KO encoded by nucleotide sequence set forth in SEQ ID NO:56 (corresponding to amino acid sequence set forth in SEQ ID NO:75)

[00208] The pre-steviol glycoside/flux, KO step/flux, and KAH step/flux values are shown in Table 12 below. In the strain comprising the KO encoded by nucleotide sequence set forth in SEQ ID NO:56 (strain A), lower accumulation of ent-kaurene, ent-kaurenol, ent-kaurnal, and ent-kaurenol glycosides resulted. Higher levels of ent-kaurenoic acid and steviol glycosides were also measured, as compared to the control strain. In the strain comprising the KAH encoded by nucleotide sequence set forth in SEQ ID NO:80, the KO encoded by nucleotide sequence set forth in SEQ ID NO:56 (corresponding to amino acid sequence set forth in SEQ ID NO:75), and the KO encoded by nucleotide sequence set forth in SEQ ID NO:65 (strain B), ent-kaurene, ent-kaurenol, ent-kaurenal, ent-kaurenol glycosides, and ent-kaurenoic acid accumulation decreased and accumulation of steviol glycosides increased, as compared to the control strain. In the strain comprising CPR12 (SEQ ID NO:97, SEQ ID NO:98), the KAH encoded by nucleotide sequence set forth in SEQ ID NO:80, and the KO encoded by nucleotide sequence set forth in SEQ ID NO:56 (strain C), ent-kaurenol, ent-kaurenal, ent-kaurenol glycosides, and ent-kaurenoic acid accumulation decreased and accumulation of steviol glycosides increased, as compared to the control. See Table 12. Thus, CPR12 was found to be a reductase protein that improves KAH and/or KO activity.

**Table 12. Pre-steviol glycoside/flux, KO step/flux, and KAH step/flux values for steviol glycoside-producing strains of Example 12.**

Strain	Pre-Steviol Glycoside/Flux	KO step/Flux	KAH step/Flux
Example 12, Strain A	0.48	0.28	0.22
Example 12, Strain B	0.64	0.18	0.12
Example 12, Strain C	0.55	0.24	0.12
Control	0.40	0.43	0.17

**[00209]** Having described the invention in detail and by reference to specific embodiments thereof, it will be apparent that modifications and variations are possible without departing from the scope of the invention defined in the appended claims. More specifically, although some aspects of the present invention are identified herein as particularly advantageous, it is contemplated that the present invention is not necessarily limited to these particular aspects of the invention.

**Table 13. Sequences disclosed herein.**

**SEQ ID NO:1**

```

MNLSLCIASP LLTKSNRPAA LSAIHTASTS HGGQTNPTNL IIDTTKERIQ KQFKNVEISV      60
SSYDTAWVAM VPSPNSPKSP CFPECLNWLI NNQLNDGSWG LVNHTHNNH PLLKDSLST      120
LACIVALKRW NVGEDQINKG LSFIESNLAS ATEKSQPSPI GFDIIFPGLL EYAKNLDINL      180
LSKQTDfSLM LHKRELEQKR CHSNEMDGYL AYISEGLGNL YDWNMVKKYQ MKNGSVFNsp      240
SATAAAFINH QNPGCLNYLN SLLDKFGNAV PTVYPHDLFI RLSMVDTIER LGISHHFRVE      300
IKNVLDETYR CWVERDEQIF MDVVTALAF RLLRINGYEV SPDPLAEITN ELALKDEYAA      360
LETYHASHIL YQEDLSSGKQ ILKSADFLKE IISTDSNRLS KLIHKEVENA LKFPINTGLE      420
RINTRRNIQL YNVDNTRILK TTYHSSNISN TDYLRLAVED FYTCQSIYRE ELKGLERWV      480
ENKLDQLKFA RQKTAYCYFS VAATLSSPEL SDARISWAKN GILTTVVDDF FDIGGTIDEL      540
TNLIQCVEKW NVDVDKDCS EHVRIILFLAL KDAICWIGDE AFKWQARDVT SHVIQTWLEL      600
MNSMLREAIW TRDAYVPTLN EYMENAYVSF ALGPVVKPAI YFVGPKLSEE IVESSEYHNL      660
FKLMSTQGRL LNDIHSFKRE FKEGKLNVA LHLNNGESGK VEEVVEEMM MMIKNKRKEL      720
MKLIFEENGs IVPRACKDAF WNMCHVLNFF YANDDGFTGN TILDTVKDII YNPLVLVNEN      780
EEQR

```

**SEQ ID NO:2**

```

MNLSLCIASP LLTKSSRPTA LSAIHTASTS HGGQTNPTNL IIDTTKERIQ KLFKNVEISV      60
SSYDTAWVAM VPSPNSPKSP CFPECLNWLI NNQLNDGSWG LVNHTHNNH PLLKDSLST      120
LACIVALKRW NVGEDQINKG LSFIESNLAS ATDKSQPSPI GFDIIFPGLL EYAKNLDINL      180
LSKQTDfSLM LHKRELEQKR CHSNEIDGYL AYISEGLGNL YDWNMVKKYQ MKNGSVFNsp      240
SATAAAFINH QNPGCLNYLN SLLDKFGNAV PTVYPLDLYI RLSMVDTIER LGISHHFRVE      300
IKNVLDETYR CWVERDEQIF MDVVTALAF RLLRIHGYKV SPDQLAEITN ELAFKDEYAA      360
LETYHASQIL YQEDLSSGKQ ILKSADFLKG ILSTDSNRLS KLIHKEVENA LKFPINTGLE      420

```

RINTRRNIQL	YNVDNTRILK	TTYHSSNISN	TYYLRLAVED	FYTCQSIYRE	ELKGLERWVV	480
QNKLDQLKFA	RQKTAYCYFS	VAATLSSPEL	SDARISWAKN	GILTTVVDDF	FDIGGTIDEL	540
TNLIQCVEKW	NVDVDKDCCS	EHVRILFLAL	KDAICWIGDE	AFKWQARDVT	SHVIQTWLEL	600
MNSMLREAIW	TRDAYVPTLN	EYMENAYVSF	ALGPIVKPAI	YFVGPKLSEE	IVESSEYHNL	660
FKLMSTQGRL	LNDIHSFKRE	FKEGKLNAVA	LHLSNGESGK	VEEEVVEEMM	MMIKNRKREL	720
MKLI FEENG	IVPRACKDAF	WNMCHVLNFF	YANDDGTGN	TILDTVKDII	YNPLVLVNEN	780
EEQR						784

## SEQ ID NO:3

MAMPVKLTPA	SLSLKAVCCR	FSSGGHALRF	GSSLPWCWRT	PTQRSTSSST	TRPAAEVSSG	60
KSKQHDQEAS	EATIROQLQL	VDVLENMGIS	RHFAAEIKCI	LDRTYRSWLQ	RHEEIMLDTM	120
TCAMAFRILR	LNGYNVSSDE	LYHVVEASGL	HNSLGGYLND	TRTLLELHKA	STVSISEDES	180
ILDSIGSRSR	TLLREQLESG	GALRKPSLFK	EVEHALDGP	YTTLDRLHHR	WNIEFNIE	240
QHMLETFYLS	NQHTSRDILA	LSIRDFSSSQ	FTYQQELQHL	ESWVKECRLD	QLQFARQKLA	300
YFYLSAAGTM	FSPELSDART	LWAKNGVLT	IVDDFFDVAG	SKEELENLVM	LVEMWDEHHK	360
VEFYSEQVEI	EIFSIYDSVN	QLGEKASLVQ	DRSITKHLVE	IWLDDLKSM	TEVEWRLSKY	420
VPTEKEYMIN	ASLIYFGLPI	VLPALYFVGP	KISESIVKDP	EYDELFLKMS	TCGRLLNDVQ	480
TFEREYNEGK	LNSVSLVLH	GGPMSISDAK	RKLQKPIDTC	RRDLLSLVLR	EESVVRPCK	540
ELFWKMCKVC	YFYSTTDG	SSQVERAKEV	DAVINEPLKL	QGSHTLVSDV		590

## SEQ ID NO:4

MSCIRPWFCP	SSISATLTDP	ASKLVTGEFK	TTSLNFGTK	ERIKKMFDKI	ELSVSSYDTA	60
WVAMVPSDC	PETPCFPECT	KWILENQLGD	GSWSLPHGNP	LLVKDALST	LACILALKRW	120
GIGEEQINKG	LRFIELNSAS	VTNEQHKPI	GFDIIFPGMI	EYAKDLNL	PLKPTDINS	180
LHRRALELTS	GGGNLEGR	AYLAYVSEGI	GKLQWEMAM	KYQRKNGSLF	NSPSTAAAF	240
IHIQDAECLH	YIRSLQKFG	NAVPTIYPLD	IYARLSMVDA	LERLGIDRHF	RKERKFVLDE	300
TYRFLWQGEE	EIFSDNATCA	LAFLRLRLNG	YDVSLEDHFS	NSLGGYKDS	GALELYRAL	360
QLSYPDESLL	EKQNSRTSYF	LKQGLSNVSL	CGDRLRKNII	GEVHDALNFP	DHANLQRLAI	420
RRRIKHYATD	DTRILKTSYR	CSTIGNQDFL	KLAVEDFNIC	QSIQREEFKH	IERWVVERRL	480
DKLKFARQKE	AYCYFSAAAT	LFAPELSDAR	MSWAKNGVLT	TVVDDFFDVG	GSEELVNLI	540
ELIERWDVNG	SADFCSEEE	IIYSAIHSTI	SEIGDKSFGW	QGRDVKSHVI	KIWLDDLKSM	600
LTEAQWSSNK	SVPTLDDEMT	TAHVSFALGP	IVLPALYFVG	PKLSEEVAGH	PELLNLYKVM	660
STCGRLNDW	RSFKRESEEG	KLNAISLYMI	HSGGASTEER	TIEHFKGLID	SQRRQLQLV	720
LQEKDSIIPR	PCKDLFWNMI	KLHFTFYMKD	DGFTSNEMRN	VVKAIINEPI	SLDEL	775

## SEQ ID NO:5

cgctcagtc	caaggcta	tcgtcgcgag	ttgctacgac	gccgtttcgg	ttgcttctgg	60
ttctcttatg	tctatcaacc	ttcgctcctc	cggttggtcg	tctccgatct	cagctacttt	120
ggaacgagga	ttggactcag	aagtacagac	aagagcta	aatgtgagct	ttgagcaaac	180
aaaggagaag	attaggaaga	tgttgagaa	agtggagctt	tctgtttcgg	cctacgatac	240
tagttgggta	gcaatggttc	catcaccgag	ctcccaaaat	gctccacttt	tcccacagtg	300
tgtgaaatgg	ttaattggata	atcaacatga	agatggatct	tggggacttg	ataaccatga	360
ccatcaatct	cttaagaagg	atgtgttatc	atctacactg	gctagtatcc	tcgcgttaaa	420
gaagtgggga	attggtgaaa	gacaaataaa	caagggtctc	cagtttattg	agctgaattc	480
tgcattagtc	actgatgaaa	ccatacagaa	accaacagg	tttgatatta	tatttcctgg	540
gatgattaaa	tatgctagag	atctggaatc	gacgattcca	ttgggctcag	aagtgggtga	600
tgacatgata	cgaaaaagag	atctggatct	taaatgtgat	agtgaagagt	tttcaaagg	660
aagagaagca	tatctggcct	atgttttaga	ggggacaaga	aacctaaaag	attgggattt	720
gatagtcaaa	tatcaaagga	aaaatgggtc	actgtttgat	tctccagcca	caacagcagc	780
tgcttttact	cagttttgga	atgatgggtg	tctccggtat	ctctgttctc	tccttcagaa	840
attcgaggct	gcagttcctt	cagtttatcc	atgtgatcaa	tatgcacgcc	ttagtataat	900
tgctactctt	gaaagcttag	gaattgatag	agattttcaa	accgaaatca	aaagcatatt	960
ggatgaaacc	tatagataat	ggcttcgtgg	ggatgaagaa	atatgtttgg	acttgccac	1020
ttgtgctttg	gctttccgat	tattgcttgc	tcatggctat	gatgtgtctt	acgatccgct	1080
aaaaccattt	gcagaagaat	ctggtttctc	tgatactttg	gaaggatatg	ttaagaatac	1140
gttttctgtg	ttagaattat	ttaaggctgc	tcaaagttat	ccacatgaat	cagctttgaa	1200
gaagcagtg	tgttggaacta	aacaatatct	ggagatggaa	ttgtccagct	gggttaagac	1260
ctctgttcga	gataaatacc	tcaagaaaga	ggtcgaggat	gctcttgctt	ttccctccta	1320
tgcaagccta	gaaagatcag	atcacaggag	aaaaatactc	aatggttctg	ctgtggaaaa	1380
caccagagtt	acaaaaacct	catatcggtt	gcacaatatt	tgacacctcg	atatcctgaa	1440
gttagctgtg	gatgacttca	atttctgcca	gtccatacac	cgtgaagaaa	tggaacgtct	1500
tgataggtgg	attgtggaga	atagattgca	ggaactgaaa	tttgccagac	agaagctggc	1560

ttactgttat	ttctctgggg	ctgcaacttt	attttctcca	gaactatctg	atgctcgtat	1620
atcgtagggc	aaaggtggag	tacttacaac	ggtttagtag	gacttctttg	atggttgagg	1680
gtccaaagaa	gaactggaag	acctcataca	cttggtcgaa	aagtgggatt	tgaacggtgt	1740
tcctgagtag	agctcagaac	atggttagat	catattctca	gttctaaggg	acaccattct	1800
cgaacacagga	gacaaagcat	tcacctatca	aggacgcaat	gtgacacacc	acattgtgaa	1860
aatgttggtg	gatctgctca	agtctatggt	gagagaagcc	gagtgggtcca	gtgacaagtc	1920
aacaccaagc	ttggaggatt	acatggaaaa	tgcgtagata	tcatttgcac	taggaccaat	1980
tgctctccca	gctacctatc	tgatcggacc	tccacttcca	gagaagacag	tcgatagcca	2040
ccaatataat	cagctctaca	agctcgtgag	cactatgggt	cgtcttctaa	atgacataca	2100
aggttttaag	agagaaagcg	cggaaaggga	gctgaatgag	gtttcattgc	acatgaaaca	2160
cgagagagac	aatcgtagca	aagaagtgat	catagaatcg	atgaaagggt	tagcagagag	2220
aaagagggaa	gaattgcata	agctagtgtt	ggaggagaaa	ggaagtgtgg	ttccaaggga	2280
atgcaaagaa	gcgttcttga	aaatgagcaa	agtgttgaa	ttattttaca	ggaaggacga	2340
tggattcaca	tcaaatgatc	tgatgagtct	tgtaaataca	gtgatctacg	agcctgttag	2400
cttagagaaa	gaatctttta	cttgatccaa	gttgatctgg	caggtaaact	cagtaaataga	2460
aaataagact	ttggtcttct	tctttgttgc	ttcagaacaa	gaagag		2506

## SEQ ID NO:6

MSINLRSSGC	SSPISATLER	GLDSEVQTRA	NNVSFEQTK	KIRKMLEKVE	LSVSAYDTSW	60
VAMVPSPPSQ	NAPLFPQCVK	WLLDNQHEDG	SWGLDNHDHQ	SLKKDVLSS	LASILALKKW	120
GIGERQINKG	LQFIELNSAL	VTDETIQKPT	GFDIIFPGMI	KYARDNLNTI	PLGSEVVDDM	180
IRKRDLDLKC	DSEKFSKGRE	AYLAYVLEGT	RNLKDWDLIV	KYQRKNGSLF	DSPATTAFAF	240
TQFGNDGCLR	YLCSLLQKFE	AAVPSVYPFD	QYARLSIIIV	LESLGIDRDF	KTEIKSILDE	300
TYRWLRGDE	EICLDLATCA	LAFRLLLAHG	YDVSVDPLKP	FAEESGFSDT	LEGYVKNTFS	360
VLELFKAAQS	YPHESALKKQ	CCWTKQYLEM	ELSSVVKTSV	RDYLLKKEVE	DALAFPSYAS	420
LERSDHRRKI	LNGSAVENTR	VTKTSYRLHN	ICTSDILKLA	VDDFNFCQSI	HREEMERLDR	480
WIVENRLQEL	KFARQKLAYC	YFSGAATLFS	PELSDARISW	AKGGVLTTVV	DDFFDVGGSK	540
EELNLIHLV	EKWDLNGVPE	YSSEHVEIIF	SVLRDTILET	GDKAFYQGR	NVTHHIVKIW	600
LDLLKSMRL	AEWSSDKSTP	SLEDYMENAY	ISFALGPVIV	PATYLGPPPL	PEKTVDSHQY	660
NQLYKLVSST	GRLLNDIQGF	KRESAEGKLN	AVSLHMKHER	DNRSEKVIIE	SMKGLAERKR	720
EELHKLVLLE	KGSVVPRECK	EAFLLKMSKVL	NLFYRKDDGF	TSNDLMSLVK	SVIYEPVSLQ	780
KESLT						785

## SEQ ID NO:7

MDAVTGLLT	PATAITIGGT	AVALAVALIF	WYLSYTSAR	RSQSNHLPRV	PEVPGVPLLG	60
NLLQLKEKKP	YMTFTRWAAT	YGPIYSIKTG	ATSMVVVSSN	EIAKEALVTR	FQSISTRNLS	120
KALKVLTADK	TMVAMSDYDD	YHKTVKRHIL	TAVLGPNAQK	KHRIHRDIMM	DNISTQLHEF	180
VKNPPEQEEV	DLRKIFQSEL	FGLAMRQALG	KDVESLYVED	LKITMNRDEI	FQVLVVDPM	240
GAIDVDWRDF	FPYLKWVPNK	KFENTIQQMY	IRREAVMKS	IKEHKKRIAS	GEKLSYIDY	300
LLSEAQTLLD	QQLMSLWEP	IESSDTTMV	TTEWAMYELA	KNPKLQDRLY	RDIVSVCGSE	360
KITKEHLSQL	PYITAIHFAD	LRRHSPVPII	PLRHVHEDTV	LGGYHVPAGT	ELAVNIYGCN	420
MDKNVWENPE	EWNPERFMKE	NETIDFQKTM	AFGGGKRVCA	GSLQALLTAS	IGIGRMVQEF	480
EWKLKDMTQE	EVNTIGLTTQ	MLRPLRAIIR	PRI			513

## SEQ ID NO:8

MAFFSMISIL	LGFVISSFIF	IFFFKLLSF	SRKNMSEVST	LPSVPPVPGF	PVIGNLLQLK	60
EKKPHKTFTR	WSEIYGPIYS	IKMGSSSLIV	LNSTETAKEA	MVTRFSSIST	RKLSNALTVL	120
TCDKSMVATS	DYDDFHKLK	RCLLNGLLGA	NAQKRKRHYR	DALIENVSSK	LHAHARDHPQ	180
EPVNFRAIFE	HELFGVALKQ	AFGKDVESIY	VKELGVTLK	DEIFKVLVHD	MMEGAIDVDW	240
RDFFPYLKW	PNKSFEARIQ	QKHKRRLAVM	NALIQDRLKQ	NGESDDDCY	LNFLMSEAKT	300
LTKEQIALIV	WETIETADT	TLVTEWAIY	ELAKHPSVQD	RLCKEIQNVC	GGEKFKEEQL	360
SQVPYLNQVF	HETLRKYSPA	PLVPYRYAHE	DTQIGGYHVP	AGSEIAINIY	GCNMDKKRWE	420
RPEDWWPERF	LDDGKYETSD	LHKTMAFGAG	KRVCAGALQA	SLMAGIAIGR	LVQEFWKLRL	480
DGEENVDTY	GLTSQKLYPL	MAINPRRS				509

## SEQ ID NO:9

MSKSNSMNST	SHETLFQQLV	LGLDRMPLMD	VHWLIYVAFG	AWLCSYVIHV	LSSSSTVKVP	60
VVGYSRVFEP	TWLLRLRFVW	EGGSIIGQGY	NKFKDSIFQV	RKLGTDIVII	PPNYIDEVRK	120
LSQDKTRSV	PFINDFAGQY	TRGMVFLQSD	LQNRVIQQL	TPKLVSILTKV	MKEELDIALT	180
KEMPEMDKND	WVEVDISSIM	VRLISRISAR	VFLGPEHCRN	QEWLTTTAEY	SESLEFITGFI	240

LRVVPILRP	FIAPLLPSYR	TLLRNVSGR	RVIGDIIRSQ	QGDGNEDILS	WMRDAATGEE	300
KQIDNIAQRM	LILSLASIHT	TAMTMTHAMY	DLACACEYIE	PLRDEVKSVV	GASGWDKTAL	360
NRPHKLDSEF	KESQRFNPVF	LLTFNRIYHQ	SMTLSDGTNI	PSGTIAVPS	HAMLQDSAHV	420
PGPTPTEFD	GFRYSKIRSD	SNYAQKYLFS	MTDSSNMAFG	YGKYACPGRF	YASNEMKLT	480
AILLLQFEFK	LPDGKGRPRN	ITIDSDMIPD	PRARLCVRKR	SLRDE		525

## SEQ ID NO:10

MEDPTVLYAC	LAIIVATFV	RWYRDPLRSI	PTVGGSDLPI	LSYIGALRWT	RRGREILQEG	60
YDGYRGSTFK	IAMLDRIWIV	ANGPKLADEV	RRRPDEELNF	MDGLGAFVQT	KYTLGEAIHN	120
DPYHVDIIRE	KLTRGLPAVL	PDVIEELTLA	VRQYIPTEGD	EWVSVNCSKA	ARDIVARASN	180
RVFVGLPACR	NQGYLDLAI	FTLSVVKDRA	IINMFPELLK	PIVGRVVGNA	TRNVRRVAVF	240
VAPLVEERRR	LMEEYGEDWS	EKPNDMLQWI	MDEAASRDSS	VKAIAERLLM	VNFAAIHTSS	300
NTITHALYHL	AEMPETLQPL	REEIEPLVKE	EGWTKAAMGK	MWWLDSFLRE	SQRYNGINIV	360
SLTRMADKDI	TLSDGTFLPK	GTLVAVPAYS	THRDDAVYAD	ALVFDPFRRS	RMRAREGEGT	420
KHQFVNTSVE	YVPFGHGKHA	CPGRFFAANE	LKAMLAYIVL	NYDVKLPDGD	KRPLNMYWGP	480
TVLPAAGQV	LFRKRQVSL					499

## SEQ ID NO:11

aaacaaagaa	tgattcaagt	tctaacaccg	atccttctct	tcctcatttt	cttcgttttc	60
tggaaggttt	acaagcacca	gaaaacccaa	atcaatcttc	caccgggaag	cttcggatgg	120
ccatttcttg	gcgaaactct	ggcactccta	cgtgcagggt	gggactcaga	gccggagaga	180
ttgttctgtg	aacggatcaa	gaaacacgga	agtcctctag	tgtttaagac	gtcgttggtt	240
ggcgaccgtt	ttgcggtggt	gtgtggacct	gccggaaaca	agttcctggt	ctgcaacgag	300
aaacagctgg	tggtcgctgt	gtggccggtt	ccggtgagga	agcttttcgg	caagtctctg	360
ctcacgattc	gtggtgatga	agctaagtgg	atgaggaaga	tggtgttatc	gtatctcggt	420
cctgatgctt	tcgcaactca	ttatgccgtc	accatggacg	tcgtcaccgg	tcggcatatc	480
gacgttccat	ggcgagggaa	ggaagaggtg	aacgtattcc	aaaccgttaa	gttatatgcc	540
tttgagcttg	catgtcggtt	attcatgaac	ctagacgacc	caaaccacat	tgcaaaactc	600
ggttccttgt	tcaacatttt	cttgaaaggc	atcattgagc	ttccaatcga	cgtcccaggg	660
acacgatttt	atagctccaa	aaaagcagca	gcagctatca	ggattgaact	aaaaaaattg	720
atataagcaa	gaaaactgga	actgaaagaa	gggaaggcat	catcttcaca	agacctctta	780
tcacatttgc	ttacatctcc	agatgaaaat	ggtatgtttc	taaccgaaga	agagattgta	840
gacaacatct	tggtactact	ctttgcgggt	catgatacct	cggctctttc	aatcactttg	900
ctcatgaaga	ctcttgccga	acattctgat	gtttatgaca	aggtgttaaa	agagcaacta	960
gagatatcga	agacgaaaga	agcatgggag	tccttgaaat	gggaggacat	acaaaagatg	1020
aaatactcct	ggagtgttat	atgtgaagtc	atgagactaa	atccacctgt	tataggaacc	1080
tatagagagg	cccttgtgga	tattgattat	gcgggttata	ccatccccaa	aggatggaag	1140
ctgcactgga	gtgctgtatc	gacacaaagg	gacgaggcta	actttgaaga	cgtaaacacgt	1200
tttgaccat	cacggtttga	aggcgcagga	ccgactccat	tcacctttgt	tccgtttgga	1260
ggggggcccta	gaatgtgttt	agggaaagaa	tttgctcgat	tggaagtact	tgcgtttctt	1320
cacaatattg	tcaccaaattt	caaattggac	ctgttgatac	ctgatgagaa	aatagaatat	1380
gatcccatgg	ctaccccgagc	aaaaggggctt	ccaattcgtc	ttcatcccca	tcaagtttga	1440
ttacttcaag	catgaatcag	tgatgtgaag	gtaaaccata	atggatctta	ttggtagtta	1500
cagattatgt	gtttttatgg	catgaagaag	ttatgataaa	taaaattgtg	ttattctaca	1560
acttatgtaa	tttgtgcctg	taagtaactg	aatctattaa	tgttttatgt	gacatgaaac	1620
ataaatgtat	aattagtaaa	ttttctgctc	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1678

## SEQ ID NO:12

MIQVLTPIIL	FLIFFVFWKV	YKHQKTKINL	PPGSFGWPFL	GETLALLRAG	WDSEPERFVR	60
ERIKKHGSPL	VFKTSLFGDR	FAVLCGPAGN	KFLFCNENKL	VASWWVPVPR	KLFGKSLITI	120
RGDEAKWMRK	MLLSYLGPD	FATHYAVTMD	VVTRRHIDVH	WRGKEEVNVE	QTVKLYAFEL	180
ACRLFMNLDD	PNHIAKLGLS	FNIFLKGIIE	LPIDVPGTRF	YSSKKAAAAI	RIELKKLIKA	240
RKLELKEGKA	SSSQDLLSHL	LTSPDENGME	LTEEEIVDNI	LLLLFAGHDT	SALSITLLMK	300
TLGEHSDVYD	KVLKEQLEIS	KTKEAWESLK	WEDIQMKYS	WSVICEVMRL	NPPVIGTYRE	360
ALVDIDYAGY	TIPKGWKLHW	SAVSTQRDEA	NFEDVTRFDP	SRFEGAGPTP	FTFVPFGGGP	420
RMCLGKEFAR	LEVLAFLHNI	VTNFKWDLII	PDEKIEYDPM	ATPAKGLPIR	LHPHQV	476

## SEQ ID NO:13

MGLFPLEDSY	ALVFEGLAIT	LALYYLLSFI	YKTSKKTCTP	PKASGEHPIT	GHLNLLSGSS	60
GLPHLALASL	ADRCGPIFTI	RLGIRRVLVV	SNWEIAKEIF	TTHDLIVSNR	PKYLAAKILG	120
FNYVSFSFAP	YGPYVVGIRK	IIATKLMSSS	RLQKLQFVRV	FELNSMKSII	RESWKEKKDE	180
EGKVLVEMKK	WFELNMNIV	LRTVAGKQYT	GTVDDADAKR	ISELFREWFH	YTGRFVVGDA	240
FPFLGWLDLG	GYKKTMEIVA	SRLDSMVSKW	LDEHRKKQAN	DDKKEDMDFM	DIMISMTEAN	300
SPLEGYGTDT	IIKTTCTMTLI	VSGVDTTSIV	LTWALSLLN	NRDTLKKAE	ELDMCVGKGR	360
QVNESDLVNL	IYLEAVLKEA	LRLYPAAFLG	GPRAFLEDCT	VAGYRIPKGT	CLLINMWKLH	420
RDPNIWSDPC	EFKPERFLTP	NQKDVDVIGM	DFELIPFGAG	RRYCPGTRLA	LQMLHIVLAT	480
LLQNFEMSTP	NDAPVDMTAS	VGMTNAKASP	LEVLLSPRVK	WS		522

## SEQ ID NO:14

MIQVLTPIII	FLIFFVFWKV	YKHQKTKINL	PPGSFGWPFL	GETLALLRAG	WDSEPERFVR	60
ERIKKHGSPL	VFKTSLFGDR	FAVLCPAGN	KFLFCNENKL	VASWWVPVPR	KLFGKSLITI	120
RGDEAKWMRK	MLLSYLGPD	FATHYAVTMD	VVTRRHIDVH	WRGKEEVNVF	QTVKLYAFEL	180
ACRLFMNLDD	PNHIAKLGSL	FNIFLKGIIIE	LPIDVPGRTRF	YSSKKAIAAI	RIELKKLIKA	240
RKLELKEGKA	SSSQDILLSH	LTSPEDEMGF	LTEEEIVDNI	LLLLFAGHDT	SALSITLLMK	300
TLGEHSDVYD	KVLKEQLEIS	KTKEAWESLK	WEDIQMKYKS	WSVICEVMRL	NPPVIGTYRE	360
ALVDIDYAGY	TIPKGWKLHW	SAVSTQRDEA	NFEDVTRFDP	SRFEGAGPTP	FTFVPPGGGP	420
RMCLGKEFAR	LEVLAFLHNI	VTNFKWDLII	PDEKIEYDPM	ATPAKGLPIR	LPHPHQV	476

## SEQ ID NO:15

MESLVVHTVN	AIWCIVIVGI	FSVGYHVVGR	AVVEQWRMRR	SLKLQGVKGP	PPSIFNGNVS	60
EMQRIQSEAK	HCSGDNIIISH	DYSSSLFPHF	DHWRKQYGR	YTYSTGLKQH	LYINHPEMVK	120
ELSQTNTLNL	GRITHITKRL	NPILNGGIIT	SNGPHWAHQ	RIIAYEFTHD	KIKGMVGLMV	180
ESAMPMLNKW	EEMVKRGGEM	GCDIRVDEDL	KDVSADVIK	ACFGSSFSKG	KAIFSMIRDL	240
LTAITKRSVL	FRFNGFTDMV	FGSKKHGDVD	IDALEMELES	SIWETVKERE	IECKDTHKKD	300
LMQLILEGAM	RSCDGNLWDK	SAYRRFVVDN	CKSIYFAGHD	STAVSVSWCL	MLLALNPSWQ	360
VKIRDEILSS	KNGIPDAES	IPNLKTVTMV	IQETMRLYPP	APIVGREASK	DIRLGLLVVP	420
KGVCITWTLIP	ALHRDPEIOW	PDANDFKPER	FSEGISKACK	YPQSYIPFGL	GPRTCVGKNF	480
GMMEVKVLVS	LIVSKFSFTL	SPTYQHSPSH	KLLVEPQHG	VIRVV		525

## SEQ ID NO:16

MYFLQYLN	TTVGVFATLF	LSYCLLLWRS	RAGNKKIAPE	AAAAWPIIGH	LHLLAGGSHQ	60
LPHITLGNMA	DKYGPVFTIR	IGLHRAVVVS	SWEMAKECST	ANDQVSSSRP	ELLASKLLGY	120
NYAMFGFSPY	GSYWREMRKI	ISLELLSNSR	LELLKDVRAS	EVVTSIKELY	KLWAEKKNES	180
GLVSVEMKQW	FGDLTLNVIL	RMVAGKRYFS	ASDASENKQA	QRCRRVFREF	FHLSGLFVVA	240
DAIPFLGWLD	WGRHEKTLKK	TAIEMDSIAQ	EWLEEHRRRK	DSGDDNSTQD	FMDVMQSVLD	300
GKNLGGYDAD	TINKATCLTL	ISGGSDDTTV	SLTWALSLVL	NNRDTLKKAQ	EELDIQVGKE	360
RLVNEQDISK	LVYLQAIVKE	TLRLYPPGPL	GGLRQFTEDC	TLGGYHVSKG	TRLIMNLSKI	420
QKDPRIWSDP	TEFQPERFLT	THKDVDPGRK	HFEFIPFGAG	RRACPGITFG	LQVLHLTLAS	480
FLHAFEFSTP	SNEQNMRES	LGLTNMKSTP	LEVLIISPRLS	SCSLYN		526

## SEQ ID NO:17

MEPNFYLSLL	LLFVTFISLS	LFFIFYKQKS	PLNLPPGKMG	YPIIGESLEF	LSTGWKGHPE	60
KFIFDRMRKY	SSELEKTSIV	GESTVVCCGA	ASNKFLFSNE	NKLVTAWWPD	SVNKIFPTTS	120
LDSNLKEESI	KMRKLLPQFF	KPEALQRYVG	VMDVIAQRHF	VTHWDNKNEI	TVYPLAKRYT	180
FLLACRLFMS	VEDENHVAKF	SDPFQLIAAG	IISLPIDLP	TPFNKAIKAS	NFIRKELIKI	240
IKQRRVDLAE	GTASPTQDIL	SHMLLTSDEN	GKSMNELNIA	DKILGLLIGG	HDTASVACTF	300
LVKYLDELPH	IYDKVYEQM	EIAKSKPAGE	LLNWDDLKMM	KYSWNVACEV	MRLSPPLQGG	360
FREAITDFMF	NGFSIPKGWK	LYWSANSTHK	NAECFFMPEK	FDPTRFEGNG	PAPYTFVPFG	420
GGPRMCPGKE	YARLEILVFM	HNLVKRFKWE	KVIPDEKIIIV	DPFPIPAKDL	PIRLYPHKA	479

## SEQ ID NO:18

atggaagcct	cttacctata	catttctatt	ttgctttttac	tggcatcata	cctgttcacc	60
actcaactta	gaaggaagag	cgctaatact	ccaccaaccg	tggttccatc	aataccaatc	120
attggacact	tatacttact	caaaaagcct	ctttatagaa	ctttagcaaa	aattgccgct	180
aagtacggac	caatactgca	attacaactc	ggctacagac	gtgttctggt	gatttcctca	240
ccatcagcag	cagaagagtg	ctttaccaat	aacgatgtaa	tcttcgcaaa	tagacctaac	300

acattgtttg	gcaaaatagt	gggtggaaca	tcccttgga	gtttatccta	cggcgatcaa	360
tggcgtaatc	taaggagagt	agcttctatc	gaaatcctat	cagttcatag	gttgaacgaa	420
tttcatgata	tcagagtggg	tgagaacaga	ttgttaatta	gaaaacttag	aagttcatct	480
tctcctgtta	ctcttataac	agtcttttat	gctctaacat	tgaacgtcat	tatgagaatg	540
atctctggca	aaagatatct	cgacagtggg	gatagagaat	tggaggagga	aggtaagaga	600
tttcgagaaa	tcttagacga	aacgttgctt	ctagccggtg	cttctaattg	tggcgactac	660
ttaccaatat	tgaactgggt	gggagttaag	tctcttgaaa	agaaattgat	cgctttgcag	720
aaaaagagag	atgacttttt	ccagggtttg	attgaacagg	ttagaaaatc	tcgtggtgct	780
aaagtaggca	aaggtagaaa	aacgatgac	gaactcttat	tatctttgca	agagtcagaa	840
cctgagtact	atacagatgc	tatgataaga	tcttttgctc	taggtctgct	ggctgcaggt	900
agtgatactt	cagcgggcac	tatggaatgg	gccatgagct	tactggtcaa	tcaccacat	960
gtattgaaga	aagctcaagc	tgaatcgat	agagttatcg	gtaataacag	attgattgac	1020
gagtcagaca	ttggaaatat	cccttacatc	gggtgtatta	tcaatgaaac	tctaagactc	1080
tatccagcag	ggccattggt	gttcccacat	gaaagttctg	ccgactgctg	tatttcggt	1140
tacaatatac	ctagaggtag	aatgttaatc	gtaaaccaat	gggcgattca	tcacgatcct	1200
aaagtctggg	atgatcctga	aacctttaaa	cctgaaagat	ttcaaggatt	agaaggaact	1260
agagatggtt	tcaaaacttat	gccattcggt	tctgggagaa	gaggatgtcc	aggtgaagg	1320
ttggcaataa	ggctgttagg	gatgacacta	ggctcagtga	tccaatgttt	tgattgggag	1380
atgagtaggag	atgagatggt	tgacatgaca	gaagggtttg	gtgtcacact	tcctaaggcc	1440
gttccattag	ttgccaaatg	taagccacgt	tccgaaatga	ctaattctcct	atccgaactt	1500
ttaa						1503

## SEQ ID NO:19

MEASYLYISI	LLLLASYLFT	TQLRRKSANL	PPTVFPSIPI	IGHLYLLKKP	LYRTLAKIAA	60
KYGPILQLQL	GYRRVLVISS	PSAABECFTN	NDVIFANRPK	TLFGKIVGGT	SLGSLSYGDQ	120
WRNLRRVASI	EILSVHRLNE	FHDIRVDENR	LLIRKLRSST	SPVTLITVFI	ALTNLVIMRM	180
ISGKRYFDSG	DRELEEEGKR	FREILDETLL	LAGASNVGDY	LPILNLWLGK	SLEKKLIALQ	240
KKRDDFFQGL	IEQVRKSRGA	KVGKGRKMTI	ELLLSLQESE	PEYYTDAMIR	SFVLGLLAAG	300
SDTSAGTMEW	AMSLLVNHPH	VLKKAQAEID	RVIGNNRLID	ESDIGNIPYI	GCIINETLRL	360
YPAGPLLPFH	ESSADCVISG	YNIPRGTMIL	VNQWAIHHPD	KVWDDPETFK	PERFQGLEGT	420
RDGFKLMPFG	SGRRGCPGEG	LAIRLLGMTL	GSVIQCFDWE	RVGDEMVDMT	EGLGVTLPKA	480
VPLVAKCKPR	SEMTNLLSEL					500

## SEQ ID NO:20

MQSDSVKVSP	FDLVSAAMNG	KAMEKLNASE	SEDPTTLPAL	KMLVENRELL	TLFTTSFAVL	60
IGCLVFLMWR	RSSSKKLVDQ	PVPQVIVVKK	KEKESEVDDG	KKKVSIFYGT	QTGTAEGFAK	120
ALVEEAKVRY	EKTSFKVIDL	DDYAADDDEY	EELKKESLA	FFFLATYGDG	EPTDNAANFY	180
KWFTEGDDKG	EWLKKLQYGV	FGLGNRQYEH	FNKIAIVVDD	KLTEMGAKRL	VPVGLGDDQ	240
CIEDDFTAWK	ELVWPELDQL	LRDEDDTSVT	TPYTAAVLEY	RVVYHDKPAD	SYAEDQHTN	300
GHVVHDAQHP	SRSNVAFKEL	LHTSQSDRSC	THLEFDISHT	GLSYETGDHV	GVYSENLSV	360
VDEALKLLGL	SPDITYFSVHA	DKEDGTPIGG	ASLPPFPFPPC	TLRDALTRYA	DVLSSPKKVA	420
LLALAAHASD	PSEADRLKFL	ASPAGKDEYA	QWIVANQRSL	LEVMSQSFPSA	KPPLGVFFAA	480
VAPRLQPRYY	SISSSPKMSP	NRIHVTCALV	YETTPAGRIH	RGLCSTWMKN	AVPLTESPDC	540
SQASIFVRTS	NRLPVPDPKV	PVIMIGPGTG	LAPFRGFLQE	RLALKESGTE	LGSSIFFFGC	600
RNRKVDFIYE	DELNNFVETG	ALSELIVAFS	REGTAKEYVQ	HKMSQKASDI	WKLLSEGAYL	660
YVCGDAKGMA	KDVHRTLHTI	VQEQQSLDSS	KAELYVKNLQ	MSGRYLRDVG		710

## SEQ ID NO:21

MTSALYASDL	FKQLKSIMGT	DSLSDDVVLV	IATTSLALVA	GFVLLWKKT	TADRSSELKP	60
LMIPKSLMAK	DEDDDLDLGS	GKTRVSIFFG	TQTGTAEGFA	KALSEEIKAR	YEKAAYKVID	120
LDDYAADDDQ	YEEKLKKETL	AFFCVATYGD	GEPTDNAARF	YKWFTEENER	DIKLQQLAYG	180
VFALGNRQYE	HFNKIGIVLD	EELCKKGAKR	LIEVGLGDDD	QSIEDDFNAW	KESLWSELDP	240
LLKDEDDKSV	ATPYTAVIPE	YRVVTHDPRF	TTQKSMESNV	ANGNTTIDIH	HPCRVDVAVQ	300
KELHTHESDR	SCIHLEFDIS	RTGITYETGD	HVGVYAENHV	EIVVEAGKLL	GHSIDLVSFI	360
HADKEDGSPL	ESAVPPFPFG	PCTLGTLGLR	YADLLNPPRK	SALVALAAYA	TEPSEAEKLL	420
HLTSPDGKDE	YSQWIVASQR	SLLEVMAAFP	SAKPLPGVFF	AAIAPRLQPR	YYSISSPRL	480
APSRVHVTS	LVYGPTPTGR	IHKGVCSTWM	KNAVPAEKSH	ECSGAPIFIR	ASNFKLPSNP	540
STPIVMVGPG	TGLAPFRGFL	QERMALKEDG	EELGSSLLFF	GCRNRQMDFI	YEDELNNFVD	600
QGVISELIMA	FSREGAQKEY	VQHKMMEKAA	QVWDLIKEEG	YLYVCGDAKG	MARDVHRTLH	660
TIVQEQQEVS	SSEAEIVKVK	LQTEGRYLRL	VW			692

## SEQ ID NO:22

MAELDTLIDIV	VLGVIFLGTV	AYFTKGKLGW	VTKDPYANGF	AAGGASKPGR	TRNIVEAMEE	60
SGKNCVVFYG	SQTGTAEDYA	SRLAKEGKSR	FGLNTMIADL	EDYDFDNLDT	VPSDNIVMFV	120
LATYGEGETP	DNAVDFYEFT	TGEDASFNEG	NDPPLGNLNY	VAFGLGNNTY	EHYNSMVRNV	180
NKALEKLGHA	RIGEAGEGDD	GAGTMEEDFL	AWKDPMEAL	AKKMGLEERE	AVYEPIFAIN	240
ERDDLTPEAN	EVYLGEPNKL	HLEGTAKGPF	NSHNPYIAP	AESYELFSAK	DRNCLHMEID	300
ISGSNLKYET	GDHIAIWPTN	PGEEVNKFLD	ILDLSGQHS	VVTVKALEPT	AKVPFPNPTT	360
YDAILRYHLE	ICAPVSRQFV	STLAAFAPND	DIKAEMNRLG	SDKDYFHEKT	GPHYNYIARF	420
LASVSKGEKW	TKIPFSAFIE	GLTKLQPRYY	SISSSSLVQP	KKISITAVVE	SQIPGRDDP	480
FRGVATNYLF	ALKQKQNGDP	NPAPFGQSYE	LTGPRNKYDG	IHVPVHVRHS	NFKLPSPDPGK	540
PIIMIGPGTG	VAPFRGFVQE	RAKQARDGVE	VGKTLFFGFC	RKSTEDFMYQ	KEWQEYKEAL	600
GDKFEMITAF	SREGSKKVYV	QHRLKERSKE	VSDLLSQKAY	FYVCGDAAHM	AREVNTVLAQ	660
IIAEGRGVSE	AKGEEIVKNM	RSANQYQVCS	DFVTLHCKET	TYANSELQED	VWS	713

## SEQ ID NO:23

atgcaatcgg	aatccgttga	agcatcgacg	attgatttga	tgactgctgt	tttgaaggac	60
acagtgatcg	atacagcgaa	cgcactctgat	aacggagact	caaagatgcc	gccggcgcttg	120
gcgatgatgt	tcgaaattcg	tgatctgttg	ctgattttga	ctacgtcagt	tgctgttttg	180
gtcggatggt	tcgttggttt	ggtgtggaag	agatcgtccg	ggaagaagtc	cggcaaggaa	240
ttggagccgc	cgaagatcgt	tgtgccgaag	aggcggctgg	agcaggaggt	tgatgatggt	300
aagaagaagg	ttacgatttt	cttcggaaca	caaactggaa	cggctgaagg	tttcgctaag	360
gcacttttcg	aagaagcgaa	agcgcgatat	gaaaaggcag	cgtttaaggt	gattgatttg	420
gatgattatg	ctgctgattt	ggatgagtat	gcagagaagc	tgaagaagga	aacatatgct	480
ttcttcttct	tggctacata	tggagatggt	gagccaactg	ataatgctgc	caaattttat	540
aaatggttta	ctgaggggaga	cgagaaaggg	gtttggcttc	aaaaacttca	atatggagta	600
tttggctctg	gcaacagaca	atatgaacat	ttcaacaaga	ttggaatagt	ggttgatgat	660
ggtctccaccg	agcagggtgc	aaaacgcatt	gttcccgttg	gtcttgaga	cgacgatcaa	720
tcaattgaag	acgatttttc	ggcatggaaa	gagtttagtg	ggcccgaatt	ggatctattg	780
cttcgcgatg	aagatgacaa	agctgctgca	actccttaca	cagctgcaat	ccctgaatac	840
cgcgctcgat	ttcatgacaa	acccgatgcg	ttttctgatg	atcatactca	aaccaatggt	900
catgctgttc	atgatgctca	acatccatgc	agatccaatg	tggctgttaa	aaaagagctt	960
catactcctg	aatccgatcg	ttcatgcaca	catcttgaat	ttgacatttc	tcacactgga	1020
ttatcttatg	aaactgggga	tcattgttgg	gtatactgtg	aaaaccctaat	tgaagttagt	1080
gaagaagctg	ggaaattggt	aggattatca	acagataact	atttctcggt	acatatgtat	1140
aacgaagatg	gttcaccact	tggtggacct	tcattacaac	ctccttttcc	tccttgtact	1200
ttagaagaaag	cattgactaa	ttatgcagat	ctgttaagct	ctcccaaaaa	gtcaactttg	1260
cttgctctag	ctgctcatgc	ttccgatccc	actgaagctg	atcgtttaag	atttcttgca	1320
tctcgcgagg	gcaaggatga	atatgctgaa	tgggttgttg	caaaccaaaag	aagtcttctt	1380
gaagtcatgg	aagcttttcc	gtcagctaga	ccgccacttg	gtgttttctt	tgacagcggt	1440
gcaccgcggt	tacagcctcg	ttactactct	atttcttctt	ccccaaagat	ggaaccaaaac	1500
aggattcatg	ttacttgccg	gttggtttat	gaaaaaactc	ccgcaggctg	tatccacaaa	1560
ggaatctgct	caacctggat	gaagaacgct	gtacctttga	cggaaagtca	agattgcagt	1620
tgggcaccga	tttttgtag	aacatcaaac	ttcagacttc	caattgacct	gaaagtcccg	1680
gttatcatga	ttggtcctg	aaccgggttg	gtccatttta	gggggtttct	tcaagaaaga	1740
ttggctctta	aagaatccgg	aaccgaactc	gggtcatcta	ttttattctt	cggttgtaga	1800
aaccgcaaag	tggattacat	atatgagaat	gaactcaaca	actttgttga	aaatggtgcg	1860
ctttctgagc	ttgatgttgc	tttctcccgc	gatggccgca	cgaagaagata	cgtgcaacat	1920
aaaatgagcc	aaaaggcttc	tgaatatagg	aatatgcttt	ctgagggagc	atatttatat	1980
gtatgttggtg	catgctaaa	catggctaaa	gatgtacacc	gtacacttca	caccatttgt	2040
caagaacagg	gaagtttgg	ctcgtctaaa	gcggagttgt	atgtgaagaa	tctacaaatg	2100
tcaggaagat	acctccgtga	tgtttggtga				2130

## SEQ ID NO:24

atgcaatccta	actccgtgaa	gatttcgccc	cttgatctgg	taactgcgct	gttttagcggc	60
aagggttttgg	acacatcgaa	cgcactcgaa	tcgggagaat	ctgctatgct	gccgactata	120
gcgatgatta	tggagaatcg	tgagctgttg	atgatactca	caacgtcggg	tgctgtattg	180
atcggatgcg	ttgtcgtttt	ggtgtggcgg	agatcgtcta	cgaagaagtc	ggcgttgagg	240
ccaccgggtga	ttgtggttcc	gaagagagtg	caagaggagg	aagttgatga	tggttaagaag	300
aaagttagcg	ttttcttcgg	caccocaaact	ggaacagctg	aaggcttcgc	taaggcactt	360
gttgaggaag	ctaaagctcg	atatgaaaag	gctgtcttta	aagtaattga	tttggtatgat	420
tatgctgctg	atgacgatga	gtatgaggag	aaactaaaga	aagaatcttt	ggcctttttc	480
tttttggtcta	cgtatggaga	tggtgagcca	acagataatg	ctgccagatt	ttataaatgg	540



tttactgagg	gagatgcgaa	aggagaatgg	cttaataaagc	ttcaatatgg	agtatttggt	600
ttgggtaaca	gacaatatga	acatttttaac	aagatcgcaa	aagtgggtga	tgatgggtctt	660
gtagaacagg	gtgcaaagcg	tcttggttcct	gttggacttg	gagatgatga	tcaatgtatt	720
gaagatgact	tcaccgcatg	gaaagagtta	gtatggccgg	agttggatca	attacttcgt	780
gatgaggatg	acacaactgt	tgctactcca	tacacagctg	ctgttgccaga	atatcgcggt	840
gtttttcatg	aaaaaccaga	cgcgctttct	gaagattata	gttatacaaa	tggccatgct	900
gttcatgatg	ctcaacatcc	atgcagatcc	aacgtggctg	tcaaaaagga	acttcatagt	960
cctgaatctg	accggtcttg	cactcatctt	gaatttgaca	tctcgaaacac	cggactatca	1020
tatgaaactg	gggaccatgt	tggagtttac	tgtgaaaact	tgagtgaagt	tgtgaatgat	1080
gctgaaaagat	tagtaggatt	accaccagac	acttactcct	ccatccacac	tgatagtga	1140
gacgggtcgc	cacttggcgg	agcctcattg	cgcctcctt	tcccgccatg	cactttaagg	1200
aaagcattga	cgtgttatgc	tgatgttttg	agttctccca	agaagtcggc	tttgcttgca	1260
ctagctgctc	atgccaccga	tcccagtgaa	gctgatagat	tgaattttct	tgcatccccc	1320
gccggaaaag	atgaatatct	tcaatggata	gttgcaagcc	aaagaagtct	ccttgaagtc	1380
atggaagcat	tcccgtcagc	taagccttca	cttggtgttt	tctttgcatc	tggtgccccg	1440
cgcttacaac	caagatacta	ctctatctt	tcctcaccca	agatggcacc	ggataggatt	1500
catgtttacat	gtgcattagt	ctatgagaaa	acacctgcag	gccgcatcca	caaaggaggt	1560
tgttcaactt	ggatgaagaa	cgcagtgctt	atgaccgaga	gtcaagattg	cagttggggc	1620
ccaatatacg	tccgaacatc	caatttcaga	ctaccatctg	accctaaggt	cccggttatc	1680
atgattggac	ctggcactgg	tttggtctct	tttagaggtt	tccttcaaga	gcggttagct	1740
ttaaagggaag	ccggaactga	cctcggttta	tccattttat	tcttcggatg	taggaatcgc	1800
aaagtggatt	tcatatatga	aaacgagctt	aacaactttg	tggagactgg	tgctctttct	1860
gagcttatgt	tygtctttct	cgtgaaaggc	ccgactaagg	aatatgtgca	acacaagatg	1920
agtgagaagg	cttcggatat	ctggaacttg	ctttctgaag	gagcatatct	atacgtatgt	1980
ggtgatgcca	aaggcatggc	caaagatgta	catcgaaacc	tccacacaat	tgtgcaagaa	2040
cagggatctc	ttgactcgtc	aaaggcagaa	ctctacgtga	agaatctaca	aatgtcagga	2100
agatacctcc	gtgacgtttg	gtaa				2124

## SEQ ID NO:25

MTSALYASDL	FKQLKSIMGT	DSLSDDVVLV	IATTSALVA	GFVLLWKKT	TADRSSELKP	60
LMIPKSLMAK	DEDDDLDLGS	GKTRVSIFFG	TQTGTAEFGA	KALSEEIKAR	YEKAAVKVID	120
LDYYAADDQ	YEEKLKKETL	AFFCVATYGD	GEPTDNAARE	YKWFTEENER	DIKLQQLAYG	180
VFALGNRQYE	HFNKIGIVLD	EELCKKGAKR	LIEVGLGDDD	QSIEDDFNAW	KESLWSELDK	240
LLKDEDDKSV	ATPYTAVIPE	YRVVTHDPRE	TTQKSMESNV	ANGNTTIDIH	HPCRVDVAVQ	300
KELHTHESDR	SCIHLEFDIS	RTGITYETGD	HVGVAENHV	EIVEEAGKLL	GHSLLDLVFSI	360
HADKEDGSPL	ESAVPPFPFG	PCTLGTGLAR	YADLLNPPRK	SALVALAAYA	TEPSEAEKLE	420
HLTSPDGKDE	YSQWIVASQR	SLLEVMAAFP	SAKPPLGVFF	AAIAPRLQPR	YYSISSSPRL	480
APSRVHVTS	LVYGPTPTGR	IHKGVCSTWM	KNAPVPAEKSH	ECSGAPIFIR	ASNFKLPSNP	540
STPIVMVGP	TGLAPFRGFL	QERMALKEDG	EELGSSLLFF	GCRNRQMDFI	YEDELNNFVD	600
QGVISELIMA	FSREGAQKEY	VQHKMMEKAA	QVWDLIKEEG	YLYVCGDAKG	MARDVHRTLH	660
TIVQEQEGVS	SSEAEAIKK	LQTEGRYLRLD	VW			692

## SEQ ID NO:26

MSSSSSSSTS	MIDLMAAIK	GEPVIVSDPA	NASAYESVAA	ELSSMLIENR	QFAMIVTTSI	60
AVLIGCIIVML	VWRRSGSGNS	KRVEPLKPLV	IKPREEEIDD	GRKKVTIFFG	TQTGTAEFGA	120
KALGEEAKAR	YEKTRFKIVD	LDYYAADDDE	YEEKLKKEDV	AFFFLATYGD	GEPTDNAARE	180
YKWFTEGNDR	GEWLKLNLYG	VFGLGNRQYE	HFNKVAKVVD	DILVEQGAQR	LVQVGLGDDD	240
QCIEDDFTAW	REALWPELDT	ILREEGDTAV	ATPYTAAVLE	YRVSIHDS	AKFNDITLAN	300
GNGYTVFDAQ	HPYKANVAVK	RELHTPESDR	SCIHLEFDIA	GSGLTMLKGD	HVGVLCDNLS	360
ETVDEALRLL	DMSPTDYFSL	HAEKEDGTPI	SSSLPPFPFP	CNLRALTALTRY	ACLLSSPKKS	420
ALVALAAHAS	DPTEAERLKH	LASPAGKDEY	SKWVVSQRS	LLEVMAEFPS	AKPPLGVFFA	480
GVAPRLQPRF	YSTSSSPKTA	ETRIHVTAL	VYEKMPTGRI	HKGVCSTWMK	NAVPEKSEK	540
LFLGRPIFVR	QSNFKLPSDS	KVPIIMIGPG	TGLAPFRGFL	QERLALVESG	VELGPSVLFF	600
GCRNRMRDFT	YEEELQRFVE	SGALAELSVA	FSREGPTKEY	VQHKMMDKAS	DIWNMISQGA	660
YLYVCGDAKG	MARDVHRSLSH	TIAQEQGSMD	STKAEGFVKV	LQTSGRYLRLD	VW	712

## SEQ ID NO:27

MQSESVEAST	IDLMTAVLKD	TVIDTANASD	NGDSKMPPAL	AMMFEIRDLL	LILTTSVAVL	60
VGCFVVLVWK	RSSGKKSKE	LEPPKIVVPK	RRLEQEVDDG	KKKVTIFFGT	QTGTAEFGAK	120
ALFEFAKARY	EKAFAFKVIDL	DDYAADLDEY	AEKLKKETYA	FFFLATYGDG	EPTDNAAKFY	180
KWFTEGDEKG	VWLQKLQYGV	FGLGNRQYEH	FNKIGIVDD	GLTEQGAARI	VPVGLGDDDQ	240
SIEDDFSAAWK	ELVWPELDDL	LRDEDDKAAA	TPYTAAIPEY	RVVFHDKPDA	FSDDHTQTNG	300

HAVHDAQHPC	RSNVAVKKEL	HTPESDRSCT	HLEFDISHTG	LSYETGDHVG	VYCENLIEVV	360
EEAGKLLGLS	TDTYFSLHID	NEDGSPLGGP	SLQPPFPFCT	LRKALTNYAD	LLSSPKKSTL	420
LALAAHASDP	TEADRLRFLA	SREGKDEYAE	WVVANQRSLL	EVMEAFPSAR	PPLGVFFAAV	480
APRLQPRYYS	ISSSPKMEPN	RIHVTCALVY	EKTPAGRIHK	GICSTWMKNA	VPLTESQDCS	540
WAPIFVRTSN	FRLPIDPKVP	VIMIGPGTGL	APFRGFLQER	LALKESGTEL	GSSILFFGCR	600
NRKVDYIYEN	ELNNFVENGA	LSELDVAFSR	DGPTKEYVQH	KMTQKASEIW	NMLSEGAYLY	660
VCGDAKGMMAK	DVHRTLHTIV	QEQGSLDSSK	AELYVKNLQM	SGRYLRDVW		709

## SEQ ID NO:28

MQNSSVKISP	LDLVTALFSG	KVLDTSNASE	SGESAMLEPTI	AMIMENRELL	MILTTSVAVL	60
IGCVVVLVWR	RSTTKSALE	PPVIVVPKRV	QEEVDDGKK	KVTVFFGTQT	GTAEGFAKAL	120
VEEAKARYEK	AVFKVIDLDD	YAADDDYEYEE	KLKKESLAFF	FLATYGDGEP	TDNAARFYKW	180
FTEGDAKGEW	LNKLQYGVFG	LGNRQYEHFN	KIAKVDDGL	VEQGAKRLLVP	VGLGDDDDQCI	240
EDDFTAWKEL	VWPELDQLLR	DEDDTTVATP	YTAAVAERYV	VFHEKPDALS	EDYSYTNNGHA	300
VHDAQHPCRS	NVAVKKELHS	PESDRSCTHL	EFDISNTGLS	YETGDHVGVI	CENLSEVVND	360
AERLVGLPPD	TYSSIHDTSE	DGSPLGGASL	PPFPFCTLR	KALTCYADVL	SSPKKSALLA	420
LAHAHATDPSE	ADRLKFLASP	AGKDEYSQWI	VASQRSLELV	MEAFPSAKPS	LGVFFASVAP	480
RLQPRYYSIS	SSPKMAPDRI	HVTCALVYEK	TPAGRIHKG	CSTWMKNAVP	MTESQDCSWA	540
PIYVRTSNFR	LPSDPKVPVI	MIGPGTGLAP	FRGFLQERLA	LKEAGTDLGL	SILFFGCRNR	600
KVDFIYENEL	NNFVETGALS	ELIVAFSREG	PTKEYVQHKM	SEKASDIWNL	LSEGAYLYVC	660
GDAKGMMAKV	HRTLHTIVQE	QGSLSKSAE	LYVKNLQMSG	RYLRDVW		707

## SEQ ID NO:29

MAEQQKIKKS	PHVLIPFPL	QGHINPFIQF	GKRLISKGVK	TTLVTTIHTL	NSTLNHSNTT	60
TTSIEIQAIS	DGCDEGGFMS	AGESYLETFK	QVGSKSLADL	IKKLQSEGTT	IDAIYDSMT	120
EWVLDVAIEF	GIDGGSFFTQ	ACVVNSLYYH	VHKGLISLPL	GETVSVPGFP	VLQRWETPLI	180
LQNHEQIQSP	WSQMLFGQFA	NIDQARWVFT	NSFYKLEEEV	IEWTRKIWNL	KVIGPTLPSM	240
YLDKRLDDDK	DNGFNLYKAN	HHECMNWLDD	KPKESVVYVA	FGSLVKHGPE	QVEEITRALI	300
DSVDNFWLWI	KHKEEGKLPE	NLSEVIKTGK	GLIVAWCKQL	DVLAHESVGC	FVTHCGFNST	360
LEAISLGPVP	VAMPQFSDQT	TNAKLLDEIL	GVGVRVKADE	NGIVRRGNLA	SCIKMIMEEE	420
RGVIIRKNAV	KWKDLAKVAV	HEGSSDNDI	VEFVSELIKA			460

## SEQ ID NO:30

MDAMATTEKK	PHVIFIPFPA	QSHIKAMLKL	AQLLHHKGLQ	ITFVNTDFIH	NQFLESSGPH	60
CLDGAPGFRF	ETIPDGVSHS	PEASIPIRES	LLRSIETNFI	DRFIDLVTKL	PDPPTCIISD	120
GFLSVFTIDA	AKKLGPVMM	YWTLAACGFM	GFYHIHSLIE	KGFAPLKDAS	YLTNGYLDTV	180
IDWVPGMEGI	RLKDFPLDWS	TDLNDKVLME	TTEAPQRSHK	VSHHIFHTFD	ELEPSIKTL	240
SLRYNHIYTI	GPLQLLLDQI	PEEKQGTGIT	SLHGYSLVKE	EPECFQWLQS	KEPNSVVYVN	300
FGSTTVMSLE	DMTEFGWGLA	NSNHFLWII	RSNLVIGENA	VLPPELEEH	KKRGFIASWC	360
SQEKVLKHPS	VGGFLTHCGW	GSTIESLSAG	VPMICWPYSW	DQLTNCRYIC	KEWEVGLEMG	420
TKVKRDEVKR	LVQELMGE	HKMRNKAKDW	KEKARIAIAP	NGSSSLNIDK	MVKEITVLAR	480
N						481

## SEQ ID NO:31

atggatgcaa	tggctacaac	tgagaagaaa	ccacacgtca	tcttcatacc	atttccagca	60
caaagccaca	ttaaagccat	gctcaaaacta	gcacaacttc	tccaccacaa	aggactccag	120
ataaccttcg	tcaacaccga	cttcatccac	aaccagtttc	ttgaatcatc	gggccccacat	180
tgtctagacg	gtgcaccggg	tttccgggttc	gaaaccattc	cggatgggtg	ttctcacagt	240
ccggaagcga	gcatcccaat	cagagaatca	ctcttgagat	ccattgaaac	caacttcttg	300
gacgttttca	ttgatcttgt	aacaaaactt	ccggatcctc	cgacttgtat	tatctcagat	360
gggttcttgt	cggttttcac	aattgacgct	gcaaaaaagc	ttggaattcc	ggtcatgatg	420
tattggacac	ttgctgcctg	tgggttcacg	gggttttacc	atattcattc	tctcattgag	480
aaaggatttg	caccacttaa	agatgcaagt	tacttgacaa	atgggtattt	ggacaccgtc	540
attgattggg	ttccgggaat	ggaaggcatc	cgtctcaagg	atttcccgtc	ggactggagc	600
actgacctca	atgacaaaag	tttgatgttc	actacggaag	ctcctcaaa	gtcacacaa	660
gtttcacatc	atattttcca	cacgttcgat	gagttggagc	ctagtattat	aaaaactttg	720
tcattgaggt	ataatcacat	ttacaccatc	ggccactgc	aattacttct	tgatcaaata	780
cccgaagaga	aaaagcaaac	tggaaattacg	agtctccatg	gatacagttt	agtaaaaagaa	840
gaaccagagt	gtttccagtg	gcttcagttc	aaagaaccaa	attccgtcgt	ttatgtaaat	900
tttggaagta	ctacagtaat	gtcttttagaa	gacatgacgg	aatttgggtg	gggacttgct	960

aatagcaacc	attatttccct	ttggatcatc	cgatcaaact	tggtgatagg	ggaaaatgca	1020
gttttgcccc	ctgaacttga	ggaacatata	aagaaaagag	gctttattgc	tagctggtgt	1080
tcacaagaaa	aggtcttgaa	gcacccttcg	gttgagggtt	tcttgactca	ttgtgggtgg	1140
ggatcgacca	tcgagagctt	gtctgctggg	gtgccaatga	tatgctggcc	ttattcgtgg	1200
gaccagctga	ccaactgtag	gtatatatgc	aaagaatggg	aggttgggct	cgagatggga	1260
accaaaagtga	aacgagatga	agtcgaagg	cttgtacaag	agttgatggg	agaaggaggt	1320
cacaaaatga	ggaacaaggc	taaagattgg	aaagaaaagg	ctcgcatgtc	aatagctcct	1380
aacggttcat	cttctttgaa	catagacaaa	atggtcaagg	aaatcacctg	gctagcaaga	1440
aactagttac	aaagtgtgtt	cacattgtgc	tttctattta	agatgtaact	ttgttctaata	1500
ttaatatgtt	ctagatgtat	tgaaccataa	gtttagtttg	tctcaggaat	tgatttttaa	1560
tgaaataatg	gtcattaggg	gtgagt				1586

## SEQ ID NO:32

atggatgcaa	tggcaactac	tgagaaaaag	cctcatgtga	tcttcattcc	atttcctgca	60
caatctcaca	taaaaggcaat	gctaaaagta	gcacaactat	tacaccataa	gggattacag	120
ataactttcg	tgaataccga	cttcatccat	aatcaatttc	tggaatctag	tggccctcat	180
tgtttgacg	gagccccagg	gtttagattc	gaaacaattc	ctgacggtgt	ttcacattcc	240
ccagaggcct	ccatcccaat	aagagagagt	ttactgaggt	caatagaaac	caactttttg	300
gatcggttca	ttgactttgt	cacaaaactt	ccagaccac	caacttgc	aatctctgat	360
ggctttctgt	cagtgtttac	tatcgacgct	gccaaaaagt	tgggtatccc	agttatgatg	420
tactggactc	ttgctgcatg	cggtttcatg	ggtttctatc	acatccattc	tcttatcgaa	480
aagggttttg	ctccactgaa	agatgcatca	tacttaacca	acggctacct	ggatactggt	540
attgactggg	taccagggtat	ggaagggtata	agacttaaaag	attttccttt	ggattggtct	600
acagacctta	atgataaaagt	attgatgttt	actacagaag	ctccacaaaag	atctcataag	660
gtttcacatc	atatctttca	cacctttgat	gaattggaac	catcaatcat	caaaaccttg	720
tctctaagat	acaatcataat	ctacactatt	ggtccattac	aattacttct	agatcaaaatt	780
cctgaagaga	aaaagcaaac	tggtattaca	tccttacacg	gctactcttt	agtgaagag	840
gaaccagaat	gttttcaatg	gctacaaaagt	aaagagccta	attctgtggt	ctacgtcaac	900
ttcggaagta	caacagtcac	gtccttgga	gatatgactg	aatttggttg	gggccttgct	960
aattcaaatc	attactttct	atggattatc	aggccaatt	tggaatagg	ggaaaacgcc	1020
gtattacctc	cagaatttga	ggaacacatc	aaaagagag	gtttcattgc	ttcctggtgt	1080
tctcaggaaa	aggtattgaa	acatccttct	gttggtggtt	tccttactca	ttgcggttgg	1140
ggctctacaa	tcgaatcact	aagtgacgga	gttccaatga	tttggtggcc	atattcatgg	1200
gaccaactta	caaattgtag	gtatatctgt	aaagagtggg	aagttggatt	agaaatggga	1260
acaaaggtta	aacgtgatga	agtgaaga	ttggttcagg	agttgatggg	ggaaggtggc	1320
cacaagatga	gaacaaggc	caaagattgg	aaggaaaaag	ccagaattgc	tattgctcct	1380
aacgggtcat	cctctctaaa	cattgataag	atggtcaaa	agattacagt	cttagccaga	1440
aactaa						1446

## SEQ ID NO:33

MKTGFISPAT	VEHHRISPAT	TFRHLSPAT	TNSTGIVALR	DINFRCKAVS	KEYSDLLQKD	60
EASFTKWDDD	KVKDHLDTNK	NLYPNDEIKE	FVESVKAMFG	SMNDGEINVS	AYDTAWVALV	120
QDVDSGSPQ	FPSSLEWIAN	NQLSDGSWD	HLLFSAHDRI	INTLACVIAL	TSWNVHPSKC	180
EKGLNFLREN	ICKLEDENAE	HMPIGFEVTF	PSLIDIAKKL	NIEVPEDTPA	LKEIYARRDI	240
KLTKIPMEVL	HKVPTTLHS	LEGMPDLEWE	KLLKLQCKDG	SFLFSPSSTA	FALMQTKDEK	300
CLQYLNTIVT	KFNGGVPNVY	PVDLFEHIWV	VDRLQRLGIA	RYFKSEIKDC	VEYINKYWTK	360
NGICWARNT	VQDIDDTAMG	FRVLRAHGYD	VTPDVFRQFE	KDGKFVCFAG	QSTQAVTGMF	420
NVYRASQMLF	PGERILEDK	KFSYNLKEK	QSTNELLDKW	IIAKDLPGEV	GYALDIPWYA	480
SLPRLETRY	LEQYGGEDDV	WIGKTLYRMG	YVSNNTYLEM	AKLDYNNYVA	VLQLEWYTIQ	540
QWYVDIGIEK	FESDNIKSVL	VSYYLAAASI	FEPERSKERI	AWAKTTILVD	KITSIFDSSQ	600
SSKEDITAFI	DKFRNKSSSK	KHSINGEPWH	EVMVALKKTL	HGFALDALMT	HSQDIHPQLH	660
QAWEMWLTKL	QDGVDTAEL	MVQMINMTAG	RWVSKELLTH	PQYQRLSTVT	NSVCHDITKL	720
HNFKENSTTV	DSKVQELVQL	VESDTPDDL	QDMQOTFLTV	MKTFYKAWC	DPNTINDHIS	780
KVFEIVI						787

## SEQ ID NO:34

MPDAHDAPPP	QIRQRTLVD	ATQLLTESAE	DAWGEVSVSE	YETARLVAHA	TWLGGHATRV	60
AFLERQHE	GSWGPPGGYR	LVPTLSAVHA	LLTCLASPAQ	DHGVPHDRLL	RAVDAGLTAL	120
RRLGTSDSP	DTIAVELVIP	SILEGIQHL	DPAHPHSRPA	FSQHRGSLVC	PGGLDGRITG	180
ALRSHAAAGT	PVPGKVWHAS	ETLGLSTEAA	SHLQPAQGI	GGSAATATW	LTRVAPSQQS	240
DSARRYLEEL	QHRYSQPVPS	ITPITYFERA	WLLNNFAAAG	VPCEAPAALL	DSLEAALTPQ	300
GAPAGAGLPP	DADDTAAVLL	ALATHGRGR	PEVLMDYRTD	GYFQCFIGER	TPSISTNAHV	360

LETLGHHVAQ	HPQDRARYGS	AMDTASAWLL	AAQKQDGSWL	DKWHASPYA	TVCCTQALAA	420
HASPATAPAR	QRAVRWVLAT	QRSDGGWGLW	HSTVEETAYA	LQILAPPSGG	GNIPVQQALT	480
RGRARLCGAL	PLTFLWHDKD	LYTPVRVVRA	ARAAALYTTR	DLLLPPL		527

## SEQ ID NO:35

MNALSEHILS	ELRRLLEMS	DGGSVGPSVY	DTAQUALRFHG	NVTGRQDAYA	WLIAQQQADG	60
GWGSADFPLF	RHAPTWAALL	ALQRADPLPG	AADAVQTATR	FLQRQPDPIA	HAVPEDAPIG	120
AELILPQFCG	EAALLGGVA	FPRHPALLPL	RQACLVKLGA	VAMLPSPGHL	LHSWEAWGTS	180
PTTACPDGDDG	SIGISPAATA	AWRAQAVTRG	STPQVGRADA	YLQMASRATR	SGIEGVFPNV	240
WPINVFEPCW	SLYTLHLAAGL	FAHPALAEAV	RVIVAQLEAR	LGVHGLGPAL	HFAADADDTA	300
VALCVLHLAG	RDPADVLRH	FEIGELFVTF	PGERNASVST	NIHALHALRL	LGKPAAGASA	360
YVEANRNPHG	LWDNEKWHVS	WLYPTAHAVA	ALAQQKQPWR	DERALAALLQ	AQRDDGGWGA	420
GRGSTFEETA	YALFALHVMD	GSEETGRRR	IAQVVARALE	WMLARHAANG	LPQTPLWIGK	480
ELYCPTRVVR	VAELAGLWLA	LRWGRRLVLA	GAGAAP			516

## SEQ ID NO:36

gacctgacca	ccaccccccg	gcgggcccctt	tcattcttttc	cttactttct	tcctctctgct	60
gctcttgccg	tttcagtgat	tattagctgc	tgtacgtgog	tgcgtacatt	gttctctctg	120
ctgacaccca	tacacgctgt	agcttctaca	cataccagtt	cgatcgcaag	ctatagcatg	180
gggcttcaat	catcgcccat	gctgctgcca	gcgcgacgg	caacggcgcc	cgccagcgcc	240
tcacagtggc	gcacggctgt	ggcggttaat	ggtaactcgt	ttatcttctt	ctacacgtaa	300
tctctattat	atacctagat	tttctccaca	ggcagatcag	attctttaca	cagctgtatt	360
ctcaaaaaaa	actcatagaa	aaaaaagaaa	aaactaaacc	aaaggagcga	cctcaacctg	420
taccagtggc	cctgctagca	gtagcttcgt	tctgtccctt	ttttttcatt	tggatcctct	480
acataaatgc	tgggtgggtg	tgtcctttca	cgcacacatc	cgcagatagc	gccccagcag	540
cattttatgtg	gggacgacgg	ctctgaaatg	aattactagt	cagtttcatg	cgtttcagtg	600
cgagtattat	agtagtagat	ctcttctccg	atatatccgg	ccaaaggag	aagagaagag	660
aaaccacaca	tctcattctc	aactagtagt	agaaaagtaa	aaacgtacta	caagcgcaag	720
cgcaaaagatg	gtttcttcat	cgtcttgcat	aacagttcct	caoctttctt	cccttgccgt	780
cgttcaacta	ggcccatgga	gttcccgcat	caagaagaag	acggatacag	tcgcgcgtccc	840
cgccggccgc	ggccggtgga	ggagggcact	ggcgcgggcc	cagcacacca	gcgaatccgc	900
cgccgtcgcc	aaaggtacgg	gtgatcgcta	gctttgatag	ctccaaatct	gagcagcaaa	960
ttaaatagct	agggttgtaa	cgcacgcacg	catgcaggtt	cgtccctaac	gcccacgtg	1020
agaaccgatg	ccgaaaagccg	ccgcacgaga	tggcctacgg	acgacgacga	cgctgagccg	1080
ctggctgacg	agatcagggc	aatgctgacg	tcgatgagcg	acggggacat	cagcgtgtcg	1140
gcgtacgaca	ccgcctgggt	gggtcttggt	cccaggctgg	acggcgcgca	gggcccgcag	1200
ttcccgcccg	ccgtgcgggtg	gatccggaac	aaccagctcc	ccgacggctc	gtggggcgac	1260
gcggccctgt	tctccgccta	cgaccgcctg	atcaacacgc	tggcgtgcgt	cgtcacgctc	1320
accaggtggg	cgctggagcc	cgagatgcgc	ggcagaggta	cgtaattact	gtgtgctggc	1380
cgatcgagag	aacacacgac	ggcagtgatc	ctcgacagaa	aacgggctgt	gctgaagact	1440
caagtgtgtg	tgtgtgtgtg	ttcacagggc	tctctttcct	cgcccggaac	atgtggaagc	1500
tagcgacgga	ggacgaggag	tccatgccga	tagggttcga	gctcgcgttc	ccttctctca	1560
tcgaactagc	caagtagctg	ggcgtccacg	acttcccgta	cgaccaccag	gctctgcagg	1620
gaatatactc	gagcagggag	atcaagatga	agaggattcc	taagggaagt	atgcacacgg	1680
ttccacatc	catttctccac	agcctggaag	ggatgcccgg	gctagactgg	gcgaagctgc	1740
tgaactgca	gtcgagcgac	gggtccttcc	tcttctctcc	cgccggccacc	cgctacgctc	1800
tcatgaacac	cggcgacgac	aggtgcttca	gctacatcga	caggacagtc	aagaaattca	1860
acggaggagg	tacgcaagca	gtagcgtaga	tacatgggca	tagcatgcat	gcatgcaatg	1920
cagcgttgcc	cactgcatgc	gccttctctc	cttctctctc	gtctcttcaa	cggttcgtct	1980
tctctcgccg	tttctcgag	tgcccaacgt	ctaccccggt	gacctttctg	agcacatatg	2040
ggctgtcgat	cgccgtggagc	gtctcgggat	ctcccgcctac	ttccagaaa	agattgagca	2100
gtgcatggac	tacgtgaaca	gcactgggac	tgaggacggg	atctgctggg	cgaggaaactc	2160
cgacgtgaag	gaggtggagc	acacggccat	ggctttccgc	ctgctacggc	tgcacggata	2220
cagcgtctcg	ccaggtacgt	aacaaacaca	aaaaaaaaaa	acgcgcagac	aacagagatc	2280
gtcacgtcat	acacacgcgt	gtcctgaaca	tttttcattt	ggtctccacc	ccatcgtagc	2340
taataataat	aaaaaaaaaa	gtgcttctgc	cctgctgtgt	tacgtgtaga	tgtgttcaag	2400
aaactcgaga	aggacgggga	gttcttcgcc	ttcgtggggc	agtcgaacca	ggcggtgacg	2460
gggatgtaca	acctcaacag	ggcctcccg	ataagcttcc	cgggggagga	cgctctgcac	2520
cgtgcagggg	ctttctcgta	cgagtttctc	aggcggaaa	aggccgaggg	agcgtctcgt	2580
gacaaatgga	tcatactctaa	ggacctgcct	ggggaggtag	tgtacacctc	ggacttccct	2640
tggtatggga	acctgcgcgc	cgtggaggcg	agagactatc	tggaacagta	cgccggcgcc	2700
gacgatgtct	ggatcgggaa	gagctcttac	aggtagatag	atcttttttag	ctattaattg	2760
gtttcagatc	gaccagataa	aatttgcatt	attggttctt	ttgatgcatg	taattgaaag	2820

ccaataaata	acctcagtat	gcgtgatggc	tgacttttgc	attggcagga	tgcctcttgt	2880
gaataacgat	gtgtatcttg	agctggctag	gatggacttc	aaccattgcc	aagccctaca	2940
tcagcttgag	tggcaaggcc	tgaaaaggta	tgtatgttac	tatatatata	cagcccgggt	3000
gttgagtttt	ttttttat	tatttttttc	gcgattacca	tttcttctcg	atgcaaaata	3060
aatctgcaca	gatcatcata	tatatccttg	atgatatata	agggcttctc	gtatatatat	3120
cttatcacct	atatatacat	aggtggtaca	ctgagaaccg	gctcatggat	ttcggagtgg	3180
cgcaagagga	tgctctgcga	gcgtatttcc	tggccgcgcg	ttccgtctac	gagccgtgcc	3240
gagccgcgga	gcggcttgcg	tgggccagag	cggcgatact	tgccaacgcc	gtctctaccc	3300
atctccgtaa	cagcccttca	ttcagagaac	gcttggaaac	ctccttgcg	tgccgcccc	3360
gtgaagaaac	ggatggatca	tggtaataag	ctgatcgatg	ggaaattaaa	aatttaagtt	3420
ttttttttct	tttttgttgc	cattatctga	gaccaatgca	atgtggtgca	tatatatcca	3480
ggttcaactc	atcaagtggg	agtacgctg	ttcttgtgaa	ggcagttctg	cggcttaccg	3540
actcgttagc	gcgagaagcg	cagccgattc	atggcgggtg	tccggaggac	atcatccaca	3600
agctactgag	atcagctgta	agttaaacgt	aacgttcaga	agaagatttt	tttttttttt	3660
tgagtttaac	aagtactacg	acatctatcg	tttttgttca	gcatgcacag	tcattcttagc	3720
tactaatacc	attattcttc	tgtgaacttg	tgtagtgggc	tgaatgggtc	agggagaagg	3780
cagatgcagc	agacagctg	tgtaatggat	ccagtgctgt	ggaacaagaa	gggtcgcgca	3840
tggttcacga	caagcaaacg	tgtctgcctt	tagctcgaat	gatcgagatc	agcgtggggc	3900
gagctgcagg	tgaggctgcg	agcgaagatg	gtgaccgtcg	gattatccag	ctcactgggt	3960
ctatatgtga	cagtctcaag	cagaagatgc	tagtatctca	ggtatagcac	atatatacta	4020
cagaaagttt	gtgcgtagtt	attatttccc	ttttttcatg	tgacgaacat	gatgacctga	4080
tgatgcatgt	atatggcttc	atataggacc	ccgagaagaa	cgaagagatg	atgagccatg	4140
tcgatgcaga	attgaagctg	cgtatacgag	agttcgttca	gtatcttctg	agactcgggt	4200
agaagaaaac	cggcagcagc	gagacaaggc	agacctttct	gagcatcggt	aaaagctggt	4260
actacgctgc	tcactgcccg	ccgcatgtgg	tagacaggca	tatttccaga	gttatttttg	4320
aaactgtttc	cgcgcgaaaa	taatggtaat	ggtagatgtg	aatgtgatat	ggagataaga	4380
gagagagaaa	atgttgatag	tggaaatttg	cgttgatgtc	gcctccacat	tccttacgca	4440
aaagtagcgt	ctgttttggg	taaaaaaat	ccagtttctg	taaattatag	aataaatcaa	4500
tcgctgtgtc	ccaaactcta	aaatgttatt	ctgtgaagta	tggaaataat	cggtcactat	4560
acctatcttg	tggatgc					4577

## SEQ ID NO:37

MVLSSSCTTV	PHLSSSLAVVQ	LGPWSSRIKK	KTDTVAVPAA	AGRWRRLAR	AQHTSESAAV	60
AKGSSLTPIV	RTDAESRRTR	WPTDDDDAEP	LVDEIRAMLT	SMSDGDIVS	AYDTAWVGLV	120
PRLDGGEQPQ	FPAAVRWIRN	NQLPDGSGWD	AALFSAYDRL	INTLACVVTL	TRWSLEPEMR	180
GRGLSFLGRN	MWKLATEDEE	SMPIGFELAF	PSLIELAKSL	GVHDFPYDHQ	ALQGIYSSRE	240
IKMKRIPEKV	MHTVPTSILH	SLEGMPGLDW	AKLLKLQSSD	GSFLFSPAAT	AYALMNTGDD	300
RCFSYIDRTV	KKFNNGVNV	YPVDLFEHIW	AVDRLERLGI	SRYFQKEIEQ	CMQVYVNRHWT	360
EDGICWARN	DVKEVDDTAM	AFRLRLHGY	SVPDVFKNF	EKDGEFFAFV	GQSNQAVTGM	420
YNLNRASQIS	FPGEDVLHRA	GAFSYEFLLR	KEAEGALRDK	WIISKDLPGE	VVYTLDFPWY	480
GMLPRVEARD	YLEQYGGGDD	VWIGKTLYRM	PLVNNDVYLE	LARMDFNHQ	ALHQLEWQGL	540
KRWYTENRLM	DFGVAQEDAL	RAYFLAAASV	YEPCAAERL	AWARAAILAN	AVSTHLRNSP	600
SFRERLEHSL	RCRPSEETDG	SWFNSSSGSD	AVLVKAVLRL	TDSLAREAQ	IHGDPEDII	660
HKLLRSAAWE	WVREKADAAD	SVCNGSSAVE	QEGSRMVHDK	QTCILLARMI	EISAGRAAGE	720
AASEDGDRRI	IQLTGSICDS	LKQKMLVSQD	PEKNEEMMSH	VDELKLRIR	EFVQYLLRLG	780
EKKTGSSETR	QTFLSIVKSC	YAAHCPPHV	VDRHISRIVF	EPVSAAK		827

## SEQ ID NO:38

cttcttctact	aaatacttag	acagagaaaa	cagagctttt	taaagccatg	tctcttctagt	60
atcatgttct	aaactccatt	ccaagtacaa	cctttctcag	ttctactaaa	acaacaatat	120
cttcttcttt	ccttaccatc	tcaggatctc	ctctcaatgt	cgctagagac	aaatccagaa	180
gcggttccat	acattgttca	aagcttcgaa	ctcaagaata	cattaattct	caagaggttc	240
aacatgattt	gcctctaata	catgagtggc	aacagcttca	aggagaagat	gtccttcaga	300
ttagtgttgg	aagtaatagt	aatgcattca	aagaagcagt	gaagagtgtg	aaaacgatct	360
tgagaaacct	aacggacggg	gaattacga	tatcggctta	cgatacagct	tgggttgcat	420
tgatcgatgc	cggagataaa	actccggcgt	ttccctccgc	cgtgaaatgg	atcgccgaga	480
accaactttc	cgatggttct	tggggagatg	cgtatctctt	ctcttatcat	gatcgcttca	540
tcaataacct	tgcctgcgtc	gttgctctaa	gatcatggaa	tctcttttct	catcaatgca	600
acaaaggaat	cacgtttttc	cgggaaaaata	ttgggaagct	agaagacgaa	aatgatgagc	660
atatgccaat	cggattcgaa	gtagcattcc	catcgttgct	tgagatagct	cgaggaataa	720
acattgatgt	accgtacgat	tctccggtct	taaaagatat	atacgccaa	aaagagctaa	780
agcttacaag	gataccaaaa	gagataatgc	acaagatacc	aacaacattg	ttgcatagtt	840
tggaggggat	gcgtgatatta	gattgggaaa	agctcttgaa	acttcaatct	caagacggat	900

ctttctctctt	ctctcctctcc	tctaccgctt	ttgcattcat	gcagaccga	gacagtaact	960
gcctcgagta	tttgcgaaat	gccgtcaaac	gtttcaatgg	aggagttccc	aatgtctttc	1020
ccgtggatct	tttcgagcac	atatggatag	tggatcggtt	acaacgttta	gggatatcga	1080
gatactttga	agaagagatt	aaagagtgtc	ttgactatgt	ccacagatat	tggaccgaca	1140
atggcatatg	ttgggctaga	tggtcccatg	tccaagacat	cgatgataca	gccatggcat	1200
ttaggctctt	aagacaacat	ggataccaag	tggtccgcaga	tgtattcaag	aactttgaga	1260
aagagggaga	gtttttctgc	tttgtggggc	aatcaaacca	agcagtaacc	ggtatgttca	1320
acctataccg	ggcatcacaa	ttggcgtttc	caagggaaga	gatattgaaa	aacgccaaag	1380
agttttctta	taattatctg	ctagaaaaac	gggagagaga	ggagttgatt	gataagtggg	1440
ttataatgaa	agacttacct	ggcgagattg	ggttttcggtt	agagattcca	tggtacgcaa	1500
gcttgccctcg	agtagagacg	agattctata	ttgatcaata	tgggtggagaa	aacgacgttt	1560
ggattggcaa	gactctttat	aggatgccat	acgtgaacaa	taatggatat	ctggaattag	1620
caaaacaaga	ttacaacaat	tgccaagctc	agcatcagct	cgaatgggac	atattccaaa	1680
agtggatga	agaaaaatag	ttaagtga	gggtgtgctg	cagaagtga	cttctcgagt	1740
gttactactt	agcggctgca	actataattt	aatcagaaag	gtcacatgag	agaatgggtt	1800
gggctaagtc	aagtgtattg	gttaaagcca	tttctctctc	ttttggggaa	tcctctgact	1860
ccagaagaag	cttctccgat	cagtttcatg	aatacattgc	caatgctcga	cgaagtgtatc	1920
atcacttta	tgacaggaac	atgagattgg	accgaccagg	atcggttcag	gccagtcggc	1980
ttgccggagt	gttaaatcggg	actttgaatc	aaatgtcttt	tgaccttttc	atgtctcatg	2040
gccgtgacgt	taacaatctc	ctctatctat	cgtggggaga	ttggatggaa	aaatggaaac	2100
tatatggaga	tgaaggagaa	ggagagctca	tggtgaagat	gataattcta	atgaagaaca	2160
atgacctaac	taacttcttc	acccacactc	acttcgttcg	tctcgcggaa	atcatcaatc	2220
gaatctgtct	tcctcgccaa	tacttaaagg	caaggagaaa	cgatgagaag	gagaagacaa	2280
taaaagagtat	ggagaaggag	atggggaaaa	tggttgagtt	agcattgtcg	gagagtgcga	2340
catttcgtga	cgtcagcatc	acgtttcttg	atgtagcaaa	agcattttac	tactttgctt	2400
tatgtggcga	tcattctccaa	actcacatct	ccaaagtctt	gtttcaaaaa	gtctagtaac	2460
ctcatcatca	tcatcgatcc	attaacaatc	agtggatcga	tgtatccata	gatgcgtgaa	2520
taatatattca	tgtagagaag	gagaacaaat	tagatcatgt	agggttatca		2570

## SEQ ID NO:39

MSLQYHVLNS	IPSTTFLSST	KTITSSSFLT	ISGSPLNVAR	DKSRSGSIHC	SKLRTQEYIN	60
SQEVQHDLPL	IHEWQQLQGE	DAPQISVGSN	SNAFKEAVKS	VKTILRNLTG	GEITISAYDT	120
AWVALIDAGD	KTPAFPSAVK	WIAENQLSDG	SWGDAYLFSY	HDRLINTLAC	VVALRSWNLF	180
PHQCNKGITF	FRENIGKLED	ENDEHMPIGF	EVAFPSLLEI	ARGINIDVPY	DSPVLKDIYA	240
KKELKLTRIP	KEIMHKIPTT	LLHSLEGMRD	LDWEKLLKLQ	SQDGSFLFSP	SSTAFAMQOT	300
RDSNLCLEYL	NAVKRFNGGV	PNVFPVDLFE	HIWIVDRLQR	LGISRYFEEE	IKECLDYVHR	360
YWTDNGICWA	RCSHVQDIDD	TAMAFRLLRQ	HGYQVSADVF	KNFEKEGEFF	CFVGQSNQAV	420
TGMFNLYRAS	QLAFPREEIL	KNAKEFSYN	LLEKREREEL	IDKWIIMKDL	PGEIGFALEI	480
PWYASLPRVE	TRFYIDQYGG	ENDVWIGKTL	YRMPYVNNNG	YLELAKQDYN	NCQAQHOLEW	540
DIFQKWYEEN	RLSEWGVRRS	ELLECYYLAA	ATIFESERSH	ERMVWAKSSV	LVKAISSSFG	600
ESSDSRRSFS	DQFHEYIANA	RRSDHHFNDR	NMRLDRPGSV	QASRLAGVLI	GTLNQMSFDL	660
FMSHGRDVNN	LLYLSWGDWM	EKWLYGDEG	EGELMVKMII	LMKNNDLTNF	FTHTHFVRLA	720
EIINRICLPR	QYLKARRNDE	KEKTIKSMK	EMGKMVELAL	SESDFTRDVS	ITFLDVAKAF	780
YYFALCGDHL	QTHISKVLQ	KV				802

## SEQ ID NO:40

MEFDEPLVDE	ARSLVQRTLQ	DYDDRYGFGT	MSCAAYDTAW	VSLVTKTVDG	RKQWLFPECF	60
EFLLETQSDA	GGWEIGNSAP	IDGILNTAAS	LLALKRHHVQT	EQIIPQHDH	KDLAAGRAERA	120
AASLRAQLAA	LDVSTTEHVG	FEIIVPAML	PLEAEDPSLV	FDFPARKPLM	KIHDAKMSRF	180
RPEYLYGKQP	MTALHSLEAF	IGKIDFDKVR	HHRTHGSMG	SPSSTAAYLM	HASQWDGDSE	240
AYLRHVIKHA	AGQGTGAVPS	AFPSTHFESS	WILTTLFRAG	FSASHLACDE	LNKLVEILEG	300
SFEKEGGAIG	YAPGFQADVD	DTAKTISTLA	VLGRDATPRQ	MIKVFEANTH	FRTYPGERDP	360
SLTANCNALS	ALLHQPDAAM	YGSQIQKITK	FVCDYWWKSD	GKIKDKWNTC	YLYPSVLLVE	420
VLVDLVSLLE	QGLKLPDVL	ELQYRVAIL	FQACLRPLLD	QDAEGSWNKS	IEATAYGILI	480
LTEARRVCFF	DRLSEPLNEA	IRRGIAFADS	MSGTEAQLNY	IWIEKVSYP	ALLTKSYLLA	540
ARWAAKSPLG	ASVGSSSLWTP	PREGLDKHVR	LFHQAELEFRS	LPEWELRASM	IEAALFTPLL	600
RAHRLDVFP	QDVGEDKYLD	VVPFFWTAAN	NRDRTYASTL	FLYDMCFIAM	LNFLQDEFME	660
ATAGILFRDH	MDDLRQLIHD	LLAEKTSPPS	SGRSSQGTGD	ADSGIEEDVS	MSDSASDSQD	720
RSPEYDLVFS	ALSTFTKHVL	QHPISQSASV	WDRKLLAREM	KAYLLAHIQQ	AEDSTPLSEL	780
KDVPQKTDVT	RVSTSTTTFF	NWVRTTSADH	ISCPYSFHFV	ACHLGAALSP	KGSNGDCYPS	840
AGEKFLAAAV	CRHLATMCRM	YNDLGSARD	SDEGNLNSLD	FPEFADSAGN	GGIEIQKAAL	900

LRLAEFERDS	YLEAFRRLLQD	ESNRVHGPGAG	GDEARLSRRR	MAILEFFAQQ	VDLYGQVYVI	960
RDISARIPKN	EVEKKRKLDD	AFN				983

## SEQ ID NO:41

MASSTLIQNR	SCGVTSSMSS	FQIFRGQPLR	FPGTRTPAAV	QCLKKRRCLR	PTESVLESSP	60
GSGSYRIVTG	PSGINPSSNG	HLQEGSLTHR	LPIPMESID	NFQSTLYVSD	IWSETLQORTE	120
CLLQVTENVQ	MNEWIEEIRM	YFRNMTLGEI	SMSPYDTAWV	ARVPALDGSH	GPQFHRSLQW	180
IIDNQLPDGD	WGEPSLFLGY	DRVCLTLACV	IALKTGWGVA	QNVERGIQFL	QSNYKMEED	240
DANHMPIGFE	IVFPAMMEDA	KALGLDLPYD	ATILQQISAE	REKKMKKIPM	AMVYKYPTTL	300
LHSLEGLHRE	VDWNKLLQLQ	SENGSFLYSP	ASTACALMYT	KDVKCFDYLN	QLLIKFDHAC	360
PNVYPVDLFE	RLWMVDRLOQ	LGISRYFERE	IRDCLQYVYR	YWKDCGIGWA	SNSSVQDVDD	420
TAMAFRLRLT	HGFDVKEDCF	RQFFKDGEFF	CFAGQSSQAV	TGMFNLSRAS	QTLFPGESLL	480
KKARTFSRNF	LRTKHENNEC	FDKWIITKDL	AGEVEYNLTF	PWYASLPRLE	HRTYLDQYGI	540
DDIWIGKSLY	KMPAVTNEVF	LKLAKADFNM	CQALHKKKELE	QVIKWNASCQ	FRDLEFARQK	600
SVECYFAGAA	TMFEPEMVQA	RLVWARCCVL	TTVLDDYFDH	GTPVEELRVF	VQAVRTWNPE	660
LINGLPEQAK	ILFMAEEAFMA	QKRQVHHHLK	HYWDKLITSA	LKEAEWAESG		720
YVPTFDEYME	VAEISVALEP	IVCSTLFFAG	HRLEDVLDLS	YDYHLVMHLV	NRVGRILNDI	780
QGMKREASQG	KISSVQIYME	EHPSVPSEAM	AIAHLQELVD	NSMQQLTYEV	LRFTAVPKSC	840
KRIHLNMAKI	MHAIFYKDTDG	FSSLTAMTGF	VKKVLFEPVP	E		881

## SEQ ID NO:42

MPGKIENGTP	KDLKTGNDFV	SAAKSLLDRA	FKSHSHSYGL	CSTSCQVYDT	AWVAMIPKTR	60
DNVQWLFPE	CFHYLLKQQA	ADGSWGSPLT	TQTAGILDTA	SAVLALLCHA	QEPLQILDVS	120
PDEMGLRIEH	GVTSLKRQLA	VWNDVEDTNH	IGVEFIIPAL	LSMLEKELDV	PSFEFFPCRSI	180
LERMHGKELG	HFDLEQVYVK	PSSLLHSLEA	FLGKLDFDRL	SHHLYHGSM	ASPSTAAAYL	240
IGATKWDEEA	EDYLRHVMRN	GAGHGNGGIS	GTFTPTTHFEC	SWIATLLKV	GFTLKQIDGD	300
GLRGLSTILL	EALRDENGVI	GFAPTADVD	DTAKALLALS	LVNQPVSPDI	MIKVFEGKDH	360
FTTFGSEDP	SLTSNLHVLL	SLLKQSNLSQ	YHPQILKTL	FTCRWWGSD	HCVKDKWNLS	420
HLPTMLLVE	AFTEVLHLID	GGELSSLFDE	SFKCKIGLSI	FQAVLRILIT	QDNDGSRGY	480
REQTCYAILA	LVQARHVCFF	THMVDRLQSC	VDRGFSWLKS	CSFHSQDLTW	TSKTAYEVGF	540
VAEAYKLAAL	QSASLEVPAA	TIGHSVTSVAV	PSSDLEKYMR	LVRKTALFSP	LDEWGLMASI	600
IESSEFFVPLL	QAQRVEIYPR	DNIKVDEDKY	LSIIPFTWVG	CNNRSRTFAS	NRWLYDMMYL	660
SLLGYQTDEY	MEAVAGPVFG	DVSLHQTID	KVIDNTMGNI	ARANGTVHSG	NGHQHESPNI	720
GQVEDTLTRF	TNSVLNHKDV	LNSSSSDQDT	LRREFRTFMH	AHITQIEDNS	RFSKQASSDA	780
FSSPEQSYFQ	WVNSTGGSHV	ACAYSFAFSN	CLMSANLLQG	KDAFPSTGQK	YLISSVMRHA	840
TNMCRMYNDF	GSAGDRAER	NVNSIHPEF	TLCNGTSQNL	DERKERLLKI	ATYEQGYLDR	900
ALEALERQSR	DDAGDRAGSK	DMRKLKIVKL	FCDVTDLYDQ	LYVIKDLSSS	MK	952

## SEQ ID NO:43

MALVNPTALF	YGTSIRTRPT	NLLNPTQKLR	PVSSSSLPSP	SSVSAILTEK	HQSNPSENNN	60
LQTHLETPFN	FDSYMLEKVN	MVNEALDASV	PLKDPIKIHE	SMRYSLLAGG	KRIRPMMCIA	120
ACEIVGGNII	NAMPAACAVE	MIHTMSLVHD	DLPVCMNDNF	RRGKPISHKV	YGEEMAVITG	180
DALLSLSEFH	IATATKGVSK	DRIVRAIGEL	ARSVGSEGLV	AGQVVDILSE	GADVGLDHLE	240
YIHIHKTAML	LESSVIGAI	MGGGSDQQIE	KLRKFARSIG	LLFQVVDIIL	DVTKSTEELG	300
KTAGKDLLTD	KTTYPKLLGI	EKSREFAEKL	NKEAQEQLSG	FDRRKAAPLI	ALANYNAYRQ	360
N						361

## SEQ ID NO:44

MAEQQISNLL	SMFDASHASQ	KLEITVQMD	TYHYRETPPD	SSSSEGGSL	RYDERRVSLP	60
LSHNAASPD	VSQLCFSTAM	SSELNHRWKS	QRLKVADSPY	NYILTLPSKG	IRGAFIDSLN	120
VWLEVPEDET	SVIKEVIGML	HNSSLIIDDF	QDNSPLRRGK	PSTHTVFGPA	QAINATATYVI	180
VKAIEKIQDI	VGHDAALADVT	GTITITIFQGG	AMDLWWTANA	IVPSIQEYLL	MVNDKTGALF	240
RLSLELLALN	SEASISDSAL	ESLSSAVSLL	GQYFQIRDDY	MNLIDNKYTD	QKGFCELDDE	300
GKYSLLTIHA	LQTDSSDLLT	NILSMRRVQG	KLTAQKRCWF	WK		342

## SEQ ID NO:45

MEKTKEKAER	ILLEPYRYLL	QLPGKQVRSK	LSQAFNHWLK	VPEDKLQIII	EVTEMLHNAS
LLIDDIEDSS	KLRRGFPVAH	SIYGVPSVIN	SANYVYFLGL	EKVLTLDHDP	AVKLFTRQLL

ELHQGGGLDI	YWRDTYCPT	EEYKAMVLQ	KTGGLFGLAV	GLMQLFSDYK	EDLKPLLDLTL	180
GLFFQIRDDY	ANLHSKEYSE	NKSFCEDLTE	GKFSFPTIHA	IWSRPESTQV	QNILRQRTEN	240
IDIKKYCVQY	LEDVGSFAYT	RHTLRELEAK	AYKQIEACGG	NPSLVALVKH	LSKMFTEENK	300

## SEQ ID NO:46

MARFYFLNAL	LMVISLQSTT	AFTPAKLAYP	TTTTALNVAS	AETSFSLDEY	LASKIGPIES	60
ALEASVKSRI	PQTDKICESM	AYSLMAGGKR	IRPVLCIAAC	EMFGGSQDVA	MPTAVALEMI	120
HTMSLIHDDL	PSMDNDDLRR	GKPTNHVVF	EDVAILAGDS	LLSTSFEHVA	RETKGVSAEK	180
IVDVIAIRLGK	SVGAEGLAGG	QVMDLECEAK	PGTTLLDLKW	IHIHKTATLL	QVAVASGAVL	240
GGATPEEVAA	CELFAMNIGL	AFQVADDILD	VTASSEDLGK	TAGKDEATDK	TTYPKLLGLE	300
ESKAYARQLI	DEAKESLAPF	GDRAAPLLAI	ADFIIDRKN			339

## SEQ ID NO:47

MHLAPRRVPR	GRRSPDRVP	ERQGALGRRR	GAGSTGCARA	AAGVHRRRG	GEADPSAAVH	60
RGWQAGGGTG	LPDEVVSTAA	ALEMFAFAL	IHDDIMDDSA	TRRGSPVHR	ALADRLGAAL	120
DPDQAGQLGV	STAILVGDLA	LTWSDELLYA	PLTPHRLAAV	LPLVTAMRAE	TVHGQYLDIT	180
SARRPGTDTS	LALRIARYKT	AAYTMERPLH	IGAALAGARP	ELLAGLSAYA	LPAGEAFQLA	240
DDLLGVFGDP	RRTGKPDLD	LRGGKHTVLV	ALAREHATPE	QRHTLDLTLG	TPGLDRQGAS	300
RLRCVLVATG	ARAEAEERLIT	ERRDQALTAL	NALTLPPLA	EALARLTIGS	TAHPA	355

## SEQ ID NO:48

MSYFDNYFNE	IVNSVNDIHK	SYISGDVPKL	YEASYHLFTS	GGKRLRPLIL	TISSDLFGGQ	60
RERAYYAGAA	IEVLHTFTLV	HDDIMQDNI	RRGLPTVHV	YGLPLAILAG	DLHAKAFQL	120
LTQALRGLPS	ETIIKAFDIF	TRSIIISEG	QAVDMEFEDR	IDIKEQEYLD	MISRKTAALF	180
SASSIGALI	AGANDNDVRL	MSDFGTNLGI	AFQIVDDILG	LTADKEKELGK	PVFSIDIREGK	240
KTILVIKTLE	LCKEDEKKIV	LKALGNKSAS	KEELMSSADI	IKKYSLDYAY	NLAEKYYKNA	300
IDSLNQVSSK	SDIPGKALKY	LAFTIRRRK				330

## SEQ ID NO:49

MVAQTFNLDT	YLSQRQQQVE	EALSAALVPA	YPERIYEAMR	YSLLAGGKRL	RPILCLAAE	60
LAGGSVEQAM	PTACALEMIH	TMSLIHDDL	AMDNDFFRRG	KPTNHKVFE	DIAILAGDAL	120
LAYAFEHIAS	QTRGVPPQLV	LQVIARIGHA	VAATGLVGGQ	VVDLESEGKA	ISLETLEYIH	180
SHKTGALLEA	SVVSGGILAG	ADEELLARLS	HYARDIGLAF	QIVDDILDVT	ATSEQLGKTA	240
GKDQAAAKAT	YPSLLGLEAS	RQKAEELIQS	AKEALRPYGS	QAEPLALAD	FITRRQH	297

## SEQ ID NO:50

MASVTLGSKI	VVHHNNHHHP	SSILTKSRSR	SCPITLTKPI	SFRSKRTVSS	SSSIVSSSVV	60
TKEDNLQSE	PSSDFMYSYI	ITKAEVLNKA	LDSAVPLREP	LKIHEAMRYS	LLAGGKRVPR	120
VLCIAACELV	GGEESTAMPA	ACAVEMIHTM	SLIHDDLPCM	DNDLRRGKP	TNHKVFGEDV	180
AVLAGDALLS	FAFEHLASAT	SSDVSPVRV	VRAVGELAKA	IGTEGLVAGQ	VVDISSEGLD	240
LNDVGLEHLE	FIHLHKTAAL	LEASAVLGAI	VGGGSDDEIE	RLRKFARCIG	LLFQVVDLIL	300
DVTKSSKELG	KTAGKDILAD	KLTPKIMGL	EKSREFAEKL	NREARDQLLG	FDSKVAPLL	360
ALANYIAYRQ	N					371

## SEQ ID NO:51

atgtcttctc	cttcctcttc	cagtacctct	atgattgatt	tgatggctgc	tattattaaa	60
ggtgaaccag	ttatcgcttc	cgaccagca	aatgcctctg	cttatgaatc	agttgctgca	120
gaattgtctt	caatgttgat	cgaaaacaga	caattcgcca	tgatcgtaac	tacatcaatc	180
gctgtttttg	tcggttgat	gtcatgttg	gtatggagaa	gatccggtag	tggttaattct	240
aaaagagtcg	aacctttgaa	accattagta	attaagccaa	gagaagaaga	aatagatgac	300
ggtagaaaaga	aagttacaat	atttttcggt	acccaaactg	gtacagctga	aggttttgca	360
aaagccttag	gtgaagaagc	taaggcaaga	tacgaaaaga	ctagattcaa	gatagtcgat	420
ttggatgact	atgccgctga	tgacgatgaa	tacgaagaaa	agttgaagaa	agaagatggt	480



gcatttttct	ttttggcaac	ctatggtgac	ggtgaaccaa	ctgacaatgc	agccagattc	540
tacaaatggt	ttacagaggg	taatgatcgt	ggtgaatggt	tgaaaaactt	aaagtacggt	600
gttttcggtt	tgggtaacag	acaatacgaa	cattttcaaca	aagttgcaaa	ggttgctcgac	660
gatatttttg	tcgaacaagg	tgctcaaaga	ttagtccaag	taggtttggg	tgacgatgac	720
caatgtatag	aagatgactt	tactgcctgg	agagaagcct	tgtggcctga	attagacaca	780
atcttgagag	aagaagggtg	caccgccgtt	gctaccccat	atactgctgc	agtattagaa	840
tacagagttt	ccatccatga	tagtgaagac	gcaaagttta	atgatatcac	tttggccaat	900
ggtaacggtt	atacagtttt	cgatgcacaa	cacccttaca	aagctaacgt	tgcatgcaag	960
agagaattac	atacaccaga	atccgacaga	agttgtatata	acttgggaatt	tgatatcgct	1020
ggttccggtt	taaccatgaa	ggtgggtgac	catgtagggtg	ttttatgcca	caatttgtct	1080
gaaactggtt	atgaagcatt	gagattgttg	gatattgtccc	ctgacactta	ttttagtgtt	1140
cacgctgaaa	aagaagatgg	tacaccaatt	tccagttctt	taccacctcc	attccctcca	1200
tgtaacttaa	gaacagcctt	gaccagatac	gcttgcttgt	tatcatcccc	taaaaagtcc	1260
gccttggttg	cttttagccgc	tcatgctagt	gactctactg	aagcagaaaag	attgaaacac	1320
ttagcatctc	cagccggtaa	agatgaatat	tcaaagtggg	tagttgaatc	tcaaagatca	1380
ttgttagaag	ttatggcaga	atttccatct	gccaagcctc	cattaggtgt	cttctttgct	1440
ggtgtagcac	ctagattgca	accaagattc	tactcaatca	gttcttcacc	taagatcgct	1500
gaaactagaa	ttcatgtttac	atgtgcatta	gtctacgaaa	agatgccaac	cggtagaatt	1560
cacaaggggt	tatgctctac	ttggatgaaa	aatgctgttc	cttacgaaaa	atcagaaaag	1620
ttgttcttag	gtagaccaat	cttcgtaaga	caatcaaaact	tcaagttgcc	ttctgattca	1680
aaggttccaa	taatcatgat	aggtcctggt	acaggttttag	ccccattcag	agggttcttg	1740
caagaaagat	tggcttttagt	tgaatctggt	gtcgaatttag	gtccttcagt	tttgttcttt	1800
ggtgttagaa	acagaagaat	ggatttctac	tatgaagaag	aattgcaaaag	attcgtcgaa	1860
tctggtgcat	tggccgaatt	atctgtagct	tttcaagag	aaggtccaac	taaggaatac	1920
gttcaacata	agatgatgga	taaggcatcc	gacatatgga	acatgatcag	tcaaggtgct	1980
tatttgtacg	tttgccgtga	cgcaaagggt	atggccagag	atgtccatag	atcctttgcac	2040
acaattgctc	aagaacaagg	ttccatggat	agtaccaaag	ctgaagggtt	cgtaaagaac	2100
ttacaaactt	ccggtagata	cttgagagat	gtctggtga			2139

## SEQ ID NO:52

atggcggaac	aacaaaagat	caagaaatca	ccacacgttc	tactcatccc	attcccttta	60
caaggccata	taaacctctt	catccagttt	ggcaaacgat	taatctccaa	agggtgtcaaa	120
acaacacttg	ttaccaccat	ccacacctta	aactcaaccc	taaacacacag	taaacaccacc	180
accacctcca	tcgaaatcca	agcaatttcc	gatggttgtg	atgaaggcgg	ttttatgagt	240
gcaggagaat	catattttga	aacattcaaa	caagttgggt	ctaaatcact	agctgactta	300
atcaagaagc	ttcaaagtga	aggaaccaca	attgatgcaa	tcatttatga	ttctatgact	360
gaatgggttt	tagatgttgc	aattgagttt	ggaatcgatg	gtggttcggt	tttcaactca	420
gcttgtgttg	taaacagctt	atattatcat	gttcataaag	gtttgatctc	tttgccattg	480
ggtgaaaactg	tttcggttcc	tggattttcca	gtgcttcaac	gggtggagac	accgttaatt	540
ttgcagaatc	atgagcaaat	acagagccct	tgtgtctcaga	tgtgttttg	tcagtgtgct	600
aatattgatc	aagcacgttg	ggtcttcaca	aatagttttt	acaagctcga	ggaagaggta	660
atagagtggg	cgagaaagat	atggaacttg	aaggtaatcg	ggccaacact	tccatccatg	720
taccttgaca	aacgacttga	tgatgataaa	gataacggat	ttaatctcta	caaagcaaac	780
catcatgagt	gcatgaactg	gttagacgat	aagccaaaag	aatcagttgt	ttacgtagca	840
tttggtagcc	tgggtgaaaca	tggacccgaa	caagtgggaag	aaatcacacg	ggctttaata	900
gatagtgatg	tcaacttctt	gtgggttatc	aaacataaaag	aagagggaaa	gctcccagaa	960
aatctttcgg	aagtaataaa	aaccggaaaag	ggtttgatgtg	tagcatggtg	caaacaattg	1020
gatgtgttag	cacacgaatc	agtaggatgc	tttgttacac	attgtgggtt	caactcaact	1080
cttgaagcaa	taagtcttgg	agtccccggt	gttgcaatgc	ctcaattttc	ggatcaaaact	1140
acaaatgcca	agcttctaga	tgaatttttg	ggtgttgag	ttagagttaa	ggctgatgag	1200
aatgggtag	tgagaagagg	aaatcttgcg	tcatgtatta	agatgattat	ggaggaggaa	1260
agaggagtaa	taatccgaaa	gaatgcggta	aaatggaaag	atttggctaa	agtagccgtt	1320
catgaagggtg	gtagctcaga	caatgatatt	gtcgaatttg	taagtgaact	aattaaggct	1380
taaaatttttg	ttgctttgta	ttttatgtgt	tatggttttt	tgatttagat	gtattcaatt	1440
aatattgaat	cataactaaa	ttcaagatta	ttgtttgtaa	tattctttgt	cctaaaattt	1500
tgcgacttaa	aaccttttagt	ttataaaaaa	aaattagaaa	atactattgc	acgga	1555

## SEQ ID NO:53

atggaaaaa	agaccgaac	aacagttaga	cgtaggcgta	gaatcattct	gtttccagta	60
ccttttcaag	ggcacatcaa	tccaatacta	caactagcca	acgttttgta	ctctaaagggt	120
ttttctatta	caatcttcoa	caccaatttc	aaacaaacca	aaacatccaa	ttaccacat	180
ttcacattca	gattcatact	tgataatgat	ccacaagatg	aacgtatttc	aaacttacct	240
acccacggtc	cttttagctgg	aatgagaatt	ccaatcatca	atgaacatgg	tgccgatgag	300

cttagaagag	aattagagtt	acttatgttg	gcattccgaag	aggacgagga	agtctcttgt	360
ctgattactg	acgctctatg	gtacttttgc	caatctgttg	ctgatatgtt	gaatttgagg	420
agattgggtac	taatgacatc	cagtctgttt	aacttttcacg	ctcatgttag	tttaccacaa	480
tttgacgaat	tggtgatactt	ggaccctgat	gacaagacta	ggtttagagga	acagggcctct	540
ggttttccta	tggtgaaagt	caaagatatc	aagtctgcct	attctaattg	gcaaattcttg	600
aaagagatct	taggaaagat	gatcaaacag	acaaaggctt	catctggagt	gatttggaac	660
agtttcaaa	agttagaaga	gtctgaattg	gagactgtaa	tcagagaaat	tccagcacct	720
tcattcctga	taccattacc	aaaacatttg	actgcttctt	cttctctttt	gttggatcat	780
gacagaacag	tttttcaatg	gttggacca	caaccaccta	gttctgtttt	gtacgtgtca	840
tttggtagta	cttctgaagt	cgatgaaaag	gacttccttg	aaatcgcaag	aggcttagtc	900
gatagtaagc	agtcattcct	ttgggtcgtg	cgctccagggt	tcgtgaaagg	ctcaacatgg	960
gtcgaaccac	ttccagatgg	ttttctaggc	gaaagaggta	gaatagtcaa	atgggttcct	1020
caacaggaag	tttttagctca	tggcgctatt	ggggcattct	ggactcattc	cggtatggaat	1080
tcaactttag	aatcagtatg	cgaaggggta	cctatgatct	tttcagattt	tggtcttgat	1140
caaccactga	acgcaagata	catgtctgat	gttttgaaa	tggtgtata	tctagaaaat	1200
ggctgggaaa	gggtgaaaat	agctaagtca	ataagacgtg	ttatggttga	tgaagagggg	1260
gagtatatca	gacaaaacgc	aagagtgtg	aagcaaaagg	ccgacgtttc	tctaataaag	1320
ggaggctctt	catacgaatc	cttagaatct	cttgtttcct	acatttcac	actgtaa	1377

## SEQ ID NO:54

MDGVDMQTI	PLRTAIAIGG	TAVALVVALY	FWFLRSYASP	SHHSNHLFPV	PEVPGVPVLG	60
NLLQLKEKKP	YMTFTKWAEM	YGPIYSIRTG	ATSMVVSSN	EIAKEVVVTR	FPSISTRKLS	120
YALKVLTEDEK	SMVAMSDYHD	YHKTVKRHIL	TAVLGPNAQK	KFRAHRDTMM	ENVSNELHAF	180
FEKNPNQEVN	LRKIFQSQLF	GLAMKQALGK	DVESIYVKDL	ETTMKREEIF	EVLVDPMMG	240
AIEVDWRDFE	PYLKWVPNKI	FENIHRMYT	RREAVMKALI	QEHKKRIASG	ENLNSYIDYL	300
LSEAQTITDK	QLLSLWEP	IESDITMT	TEWAMYELAK	NPNMQDRLYE	EIQSVCGSEK	360
ITEENLSQLP	YLYAVFQETL	RKHCPVPIMP	LRYVHENTVL	GGYHVPAGTE	VAINIYGCNM	420
DKKVWENPEE	WNPERFLSEK	ESMDLYKTMA	FGGKRVKAG	SLQAMVISC	GIGRLVQDFE	480
WKLKDDAEED	VNTLGLTTQK	LHPLALINP	RK			512

## SEQ ID NO:55

aagcttacta	gtaaaatgga	cggtgtcatc	gatattgcaa	ccattccatt	gagaaccgct	60
attgtctattg	gtgtgtactgc	tggtgtcttg	gttgttgc	tatacttttg	gttcttgaga	120
tcctacgctt	ccccatctca	tcattctaat	catttgccac	cagtacctga	agttccagggt	180
gttccagttt	tggttaattt	gttgcaattg	aaagaaaaaa	agccttacat	gaccttcacc	240
aatgtggctg	aaatgtatgg	tcctaatctac	tctattagaa	ctggtgtctac	ttccatggtt	300
gttgtctctt	ctaaccgaaat	cgccaaagaa	gttgttgtta	ccagattccc	atctatctct	360
accagaaaat	tgtcttacgc	cttgaagggt	ttgaccgaag	ataagtctat	ggttgccatg	420
tctgattatc	acgattacca	taagaccgtc	aagagacata	ttttgactgc	tggtttgggt	480
ccaaacgccc	aaaaaaagt	tagagcacat	agagacacca	tgatggaaaa	cgtttccaat	540
gaattgcatg	ccttcttcga	aaagaaccca	aatcaagaag	tcaacttgag	aaagatcttc	600
caatcccaat	tattcggttt	ggctatgaag	caagccttgg	gtaaagatgt	tgaatccatc	660
tacgttaagg	atttggaac	caccatgaag	agagaagaaa	tcttcgaagt	tttggttgtc	720
gatccaatga	tggtgtctat	tgaagtgtat	tgagagagact	ttttcccata	cttgaaatgg	780
gttccaaaaca	agtccttcga	aaacatcatc	catagaatgt	acactagaag	agaagctggt	840
atgaaggcct	tgatccaaga	acacaagaaa	agaattgcct	ccggtgaaaa	cttgaactcc	900
tacattgatt	actgtttgtc	tgaagcccaa	accttgaccg	ataagcaatt	attgatgtct	960
ttgtgggaac	ctattatcga	atctctctgt	accactatgg	ttactactga	atgggtctatg	1020
tacgaattgc	ctaagaatcc	aaacatgcaa	gacagattat	acgaagaaat	ccaatccgtt	1080
tgcggttcg	aaaagattac	tgaagaaaaa	ttgtcccaat	tgccatactt	gtacgtgtt	1140
ttccaagaaa	ctttgagaaa	gcactgtcca	gttcctatta	tgccattgag	atatgttcac	1200
gaaaacacgc	ttttgggtgg	ttatcatgtt	ccagctggta	ctgaagttgc	tattaacatc	1260
tacggttgca	acatggataa	gaaggtctgg	gaaaatccag	aagaatggaa	tccagaaaga	1320
ttcttgtcgc	aaaaagaatc	catggacttg	tacaaaacta	tggtttttgg	tggtggtaaa	1380
agagtttgcg	ctggttcttt	acaagccatg	gttatctctt	gcattggtat	cggtagattg	1440
gtccaagatt	ttgaatggaa	gttgaaggat	gatgccgaag	aagatgttaa	cactttgggt	1500
ttgactaccc	aaaagttgca	tccattattg	gccttgatta	acccaagaaa	gtaactcgag	1560
ccgcgg						1566

## SEQ ID NO:56

atggccaccc	tccttgagca	tttccaaagt	atgccccttg	ccatccctat	tgactgggt	60
gctctgtctt	ggctgttcct	cttttacatc	aaagtctcat	tcttttccaa	caagagtgtc	120

caggctaagc	tccctcctgt	gccagtggtt	cctgggctgc	cggtgattgg	gaatttactg	180
caactcaagg	agaagaaacc	ctaccagact	tttacaaggt	gggctgagga	gtatggacca	240
atctatttcta	tcaggactgg	tgcttccacc	atggctggtc	tcaataccac	ccaagttgca	300
aaagaggcca	tggtagaccag	atatattatcc	atctcaacca	gaaagctatc	aaacgcacta	360
aagattctta	ctgctgataa	atgtatgggt	gcaataagtg	actacaacga	ttttcacaag	420
atgataaagc	gatacatact	ctcaaagtgt	cctggaccta	gtgctcagaa	gcgtcaccgg	480
agcaacagag	ataccttgag	agctaattgtc	tgacgccgat	tgcatcttca	agtaaagaac	540
tctcctcgag	aagctgtgaa	tttcagaaga	gtttttgagt	gggaactctt	tggaaattgca	600
ttgaagcaag	ccttttgaaa	ggacatagaa	aagcccattt	atgtggagga	acttggcact	660
acactgtcaa	gagatgagat	ctttaagggt	ctagtgcctg	acataatgga	gggtgcaatt	720
gaggttgatt	ggagagattt	cttcccttac	ctgagatgga	ttccgaatac	gcgcattgga	780
acaaaaattc	agcgactcta	tttcgcagg	aaagcagtga	tgactgccct	gatcaacgag	840
cagaagaagc	gaattgcttc	aggagaggaa	atcaactgtt	atatcgactt	cttgcttaag	900
gaaggggaaga	cactgacaat	ggaccaaata	agtatgttgc	tttgggagac	ggttattgaa	960
acagcagata	ctacaatggt	aacgcacagaa	tgggctatgt	atgaagttgc	taaagactca	1020
aagcgtcagg	atcgtctcta	tcaggaaatc	caaaagggtt	gtggatcgga	gatggttaca	1080
gaggaatact	tgtcccaact	gccgtacctg	aatgcagttt	tccatgaaac	gctaagggaag	1140
cacagtcagg	ctgctgttagt	tcctttaaga	tatgcacatg	aagataccca	actaggaggt	1200
tactacattc	cagctggaa	tgagattgct	ataaacatat	acgggtgtaa	catggacaag	1260
catcaatggg	aaagccctga	ggaatggaaa	ccggagagat	ttttggaccc	gaaatttgat	1320
cctatggatt	ctatggcttt	ggggctggaa	agagggtaga	tgctgggtct	1380	
cttcaggcaa	tgttaatagc	gtgcccagc	attggtaggc	tggtgcagga	gtttgagtgg	1440
aagctgagag	atggagaaga	agaaaatgta	gatactgttg	ggctcaccac	tcacaaacgc	1500
tatccaatgc	atgcaatcct	gaagccaaga	agtta			1535

## SEQ ID NO:57

aagcttacta	gtaaaatggc	ctccatcacc	catttcttac	aagattttca	agctactcca	60
ttcgctactg	cttttgctgt	tgggtggtgt	tctttgttga	tattcttctt	cttcatccgt	120
ggtttccact	ctactaagaa	aaacgaatat	tacaagttgc	caccagttcc	agttgttcca	180
ggtttgccag	tgttggtgaa	tttgttgcaa	ttgaaagaaa	agaagccata	caagactttc	240
ttgagatggg	ctgaaattca	tgttccaatc	tactctatta	gaactgggtc	ttctaccatg	300
gttgttggtta	actctactca	tgttgcctaaa	gaagctatgg	ttaccagatt	ctcttcaatc	360
tctaccagaa	agttgtccaa	ggctttggaa	ttattgacct	ccaacaaatc	tatggttgcc	420
acctctgatt	acaacgaatt	tcacaagatg	gtcaagaagt	acatcttggc	cgaattattg	480
gggtgctaag	ctcaaaagag	acacagaatt	catagagaca	ccttgatcga	aaacgtcttg	540
aacaaattgc	atgcccatat	caagaattct	ccattgcaag	ctgttaactt	cagaaagatc	600
ttcgaatctg	aattattcgg	tttggctatg	aagcaagcct	tgggttatga	tgttgattcc	660
ttgttcgttg	aagaattggg	tactaccttg	tccagagaag	aaatctacaa	cgtttttggtc	720
agtgcacatg	tgaagggtgc	tattgaagtt	gattggagag	actttttccc	atacttgaaa	780
tggtatcccaa	acaagtcctt	cgaatgaag	attcaaaagt	tggcctctag	aagacaagcc	840
gttatgaaat	ctattgtcaa	agaacaaaag	aagtcatttg	cctctggtta	gggtgaaaac	900
tgttacttga	attacttggt	gtccgaagct	aagactttga	ccgaaaagca	aatttccatt	960
ttggcctggg	aaaccattat	tgaactgct	gatacaactg	ttgttaccac	tgaatgggct	1020
atgtacgaat	tggctaaaaa	cccaaaagcaa	caagacagat	tatacaacga	aatccaaaac	1080
gtctgcggta	ctgataagat	taccgaagaa	catttgtcca	agttgcctta	cttgtctgct	1140
gtttttcacg	aaaccttgag	aaagtattct	ccatctccat	tggttccatt	gagatacgct	1200
catgaagata	ctcaattggg	tggttattat	gttccagccg	gtactgaaat	tgtgtttaat	1260
atctacgggt	gcaacatgga	caagaatcaa	tgggaaactc	cagaagaatg	gaagccagaa	1320
agatttttgg	acgaaaagta	cgatccaatg	gacatgtaca	agactatgtc	ttttggttcc	1380
ggtaaaaagag	tttgcgctgg	ttctttacaa	gctagtgttg	ttgcttgtag	ctccatcggt	1440
agattgggttc	aagaatttga	atggagattg	aaagacgggtg	aagttgaaaa	cgttgatacc	1500
ttgggtttga	ctaccataaa	gttgtatcca	atgcaagcta	tcttgcaacc	tagaaactga	1560
ctcgagccgc	gg					1572

## SEQ ID NO:58

atgatttccct	tgttggtggg	ttttgttgct	tctcctctct	tgtttatctt	cttcttgaaa	60
aaattgttgt	tcttcttcag	tcgtcacaaa	atgtccgaag	tttctagatt	gccatctgtt	120
ccagttccag	gttttccatt	gattggtaac	ttgttgcaat	tgaagaaaaa	gaagccacac	180
aagactttca	ccaagtggtc	tgaattatat	ggtccaatct	actctatcaa	gatgggttcc	240
tcttctttga	tcgtcttgaa	ctctattgaa	accgccaaag	aagctatggt	cagtagattc	300
tcttcaatct	ctaccagaaa	gttgtctaac	gctttgactg	ttttgacctg	caacaaatct	360
atgggttgcta	cctctgatta	cgatgacttt	cataagttcg	tcaagagatg	cttgttgaa	420
ggtttgggtg	gtgctaattgc	tcaagaaaga	aaaagacatt	acagagatgc	cttgatcgaa	480

aacgttacct	ctaaattgca	tgcccatacc	agaaatcatc	cacaagaacc	agttaacttc	540
agagccattt	tcgaacacga	attattcggg	gttgctttga	aacaagcctt	cggtaaagat	600
gtcgaatcca	tctatgtaaa	agaattgggt	gtcaccttgt	ccagagatga	aattttcaag	660
gtttttgggt	acgacatgat	ggaagggtgt	attgatgttg	attggagaga	tttcttccca	720
tacttgaaat	ggatcccaaa	caactctttc	gaagccagaa	ttcaacaaaa	gcacaagaga	780
agattggctg	ttatgaacgc	cttgatccaa	gacagattga	atcaaaacga	ttccgaatcc	840
gatgatgact	gctacttgaa	tttcttgatg	tctgaagcta	agaccttgac	catggaacaa	900
attgctattt	tggtttggga	aaccattatc	gaaactgctg	ataccacttt	ggttactact	960
gaatgggcta	tgtacgaatt	ggccaaacat	caatctgttc	aagatagatt	attcaaaaga	1020
atccaatccg	tctgcggttg	tgaaaagatc	aaagaagaac	aattgccaag	attgccttac	1080
gtcaatgggt	tttttcacga	aaccttgaga	aagtattctc	cagctccatt	ggttccaatt	1140
agatacgctc	atgaagatac	ccaaattggg	ggttatcata	ttccagccgg	ttctgaaatt	1200
gccatttaaca	tctacggttg	caacatggat	aagaagagat	gggaaaagacc	tgaagaatgg	1260
tgccagaaaa	gatttttggg	agatagatac	gaatcctccg	acttgcataa	gactatggct	1320
tttggtgctg	gtaaaagagt	ttgtgctggt	gctttacaag	ctagtgtgat	ggctggtatt	1380
gctatcggtg	gattggttca	agaattcgaa	tggaagttga	gagatggtga	agaagaaaac	1440
gttgatactt	acggtttgac	ctcccaaaag	ttgtatccat	tgatggccat	tatcaaccac	1500
agaagatctt	aa					1512

## SEQ ID NO:59

atggatgctg	tgacggggtt	gttaactgtc	ccagcaaccg	ctataactat	tggtggaact	60
gctgtagcat	tgccggtagc	gctaattctt	tggtacctga	aatcctacac	atcagctaga	120
agatcccaat	caaatcatct	tccaagagtg	cctgaagtcc	caggtgttcc	attgttagga	180
aatctgtttac	aattgaagga	tacatgactt	ttacgagatg	ggcagcgaca		240
tatggaccta	tctatagtat	caaaactggg	gctacaagta	tggttgtggt	atcatctaatt	300
gagatagcca	aggaggcatt	ggtgaccaga	ttccaatcca	tatctacaag	gaacttatct	360
aaagccctga	aagtacttac	agcagataag	acaatggctg	caatgtcaga	ttatgatgat	420
tatcataaaa	cagttaagag	acacatactg	accgcgtctt	tggttcctaa	tgccacagaaa	480
aagcatagaa	ttcacagaga	tatcatgatg	gataacatat	ctactcaact	tcatgaattc	540
gtgaaaaaca	accagaaca	ggaagaggta	gaccttagaa	aaatctttca	atctgagtta	600
ttcggccttag	ctatgagaca	agccttagga	aaggatgttg	aaagtgttga	cgttgaagac	660
ctgaaaatca	ctatgaatag	agacgaaatc	tttcaagtcc	ttgttgttga	tccaatgatg	720
ggagcaatcg	atgttgattg	gagagacttc	tttccatacc	taaagtgggt	cccaaacaaa	780
aagttcgaaa	atactattca	acaaatgtac	atcagaagag	aagctgttat	gaaatcttta	840
atcaaagagc	acaaaaagag	aatagcgtca	ggcgaaaagc	taaatagtta	tatcgattac	900
cttttatctg	aagctcaaac	tttaaccgat	cagcaactat	tgatgtcctt	gtgggaacca	960
atcattgaat	cttcagatac	aacaatgggtc	acaacagaat	gggcaatgta	cgaattagct	1020
aaaaacccta	aattgcaaga	taggttgtac	agagacatta	agtccgtctg	tggtatctgaa	1080
aagataaccg	aagagcatct	atcacagctg	ccttacatta	cagctatttt	ccacgaaaca	1140
ctgagaagac	actcaccagt	tcctatcatt	cctctaagac	atgtacatga	agataccggt	1200
ctaggcggtc	accatgtttc	tgctggcaca	gaacttgccg	ttaacatcta	cggttgcaac	1260
atggacaaaa	acgttttggg	aaatccagag	gaatggaaac	cagaaagatt	catgaaagag	1320
aatgagacaa	ttgatatttca	aaagacgatg	gccttcggtg	gtggtaagag	agtttgtgct	1380
ggttccttgc	aagccctttt	aactgcattc	attgggattg	ggagaatggt	tcaagagttc	1440
gaatggaaac	tgaaggatat	gactcaagag	gaagtgaaca	cgataggcct	aactacacaa	1500
atgttaagac	cattgagagc	tattatcaaa	cctaggatct	aa		1542

## SEQ ID NO:60

aagcttacta	gtaaaatgga	catgatgggt	attgaagctg	ttccatttgc	tactgctggt	60
gtttttgggtg	gtatttcctt	ggttggtttg	atcttcatca	gaagattcgt	ttccaacaga	120
aagagatccg	ttgaaggttt	gccaccagtt	ccagatattc	caggtttacc	attgattggg	180
aactttgttg	aattgaaaaga	aaagaagcca	cataagacct	ttgctagatg	ggctgaaact	240
tacggtccaa	ttttctctat	tagaactggt	gcttctacca	tgatcgtctt	gaattcttct	300
gaagttgcc	aagaagctat	ggtcactaga	ttctcttcaa	tctctaccag	aaagttgtcc	360
aacgccttga	agattttgac	cttcgataag	tgtatgggtg	ccacctctga	ttacaacgat	420
tttcacaaaa	tggtcaaggg	tttcatcttg	agaaacgttt	taggtgctcc	agcccaaaaa	480
agacatagat	gtcatagaga	taccttgatc	gaaaacatct	ctaagtactt	gcatgcccat	540
gttaagactt	ctccatttga	accagttgtc	ttgaagaaga	ttttcgaatc	cgaaattttc	600
ggttttgctt	tgaacaaagc	cttgggtaag	gatatcgaat	ccatctatgt	tgaagaattg	660
ggtactacct	tgtccagaga	agaaattttt	gccgttttgg	ttgttgatcc	aatggctggg	720
gctattgaag	ttgattggag	agattttttc	ccatacttgt	cctggattcc	aaacaagtct	780

atggaaatga	agatccaaa	aatggatttt	agaagaggtg	ctttgatgaa	ggccttgatt	840
ggtgaacaaa	agaaaaaat	cgggtccggg	gaagaaaaga	actcctacat	tgatttcttg	900
ttgtctgaag	ctaccacttt	gaccgaaaag	caaattgcta	tgttgatctg	ggaaaccatc	960
atcgaaatgt	ccgatacaac	tttgggtacc	tctgaatggg	ctatgtacga	attggctaaa	1020
gacccaaaata	gacaagaaat	cttgtacaga	gaaatccaca	aggttttcgg	ttctaacaag	1080
ttgactgaag	aaaacttgtc	caagttgcca	tacttgaact	ctgttttcca	cgaaaccttg	1140
agaaagtatt	ctccagctcc	aatggttcca	gttagatatg	ctcatgaaga	tactcaattg	1200
ggtggttacc	atattccagc	tggttctcaa	attgccatta	acatctacgg	ttgcaacatg	1260
aacaaaaagc	aatgggaaaa	tcttgaagaa	tggaagccag	aaagattctt	ggacgaaaa	1320
tatgacttga	tggacttgca	taagactatg	gcttttgggt	gtggtaaaag	agtgttgctt	1380
ggtgctttac	agaacatggt	gattgcttgc	acttccatcg	gtagattcgt	tcaagaatgt	1440
gaatgggaagt	tgtggtgtgg	tgaagaagaa	aacgttgata	ctgttgcttt	gacctccaa	1500
aaattgcatc	caatgcaagc	cattatttaag	gccagagaat	gactcgagcc	gcgg	1554

## SEQ ID NO:61

atgcaatcag	attcagtcac	agtctctcca	tttgatttgg	tttccgctgc	tatgaatggc	60
aaggcaatgg	aaaagtggaa	cgctagtga	tctgaagatc	caacaacatt	gcctgcacta	120
aagatgctag	ttgaaaatag	agaattgttg	acactgttca	caacttcctt	cgcagttctt	180
attgggtgtc	ttgtatttct	aatgtggaga	cgttcatcct	ctaaaaagct	ggtacaagat	240
ccagttccac	aagttatcgt	tgtaaaaga	aaagagaagg	agtcagaggt	tgatgacggg	300
aaaaagaaa	tttctatttt	ctacggcaca	caaacaggaa	ctgccgaagg	ttttgctaaa	360
gcattagtcg	aggaagcaaa	agtgaatat	gaaaaagcct	ctttcaaggt	tatcgatcta	420
gatgactacg	ctgcagatga	tgatgaatat	gaggaaaaac	tgaaaaagga	atccttagcc	480
ttcttcttct	tggccacata	cggatgatgt	gaacctactg	ataatgctgc	taacttctac	540
aagtggttca	cagaaggcga	cgataaaggt	gaatggctga	aaaagttaca	atcggagta	600
tttggtttag	gtaacagaca	atatgaacat	ttcaacaaga	tcgctattgt	agttgatgat	660
aaacttactg	aaatgggagc	caaaaagatta	gtaccagtag	gattagggga	tgatgatcag	720
tgatataga	atgacttcac	cgcctggaag	gaattggtat	ggccagaatt	ggatcaactt	780
ttaaggagac	aagatgatac	ttctgtgact	accccataca	ctgcagccgt	attggagtac	840
agagtgggtt	accatgataa	accagcagac	tcatatgctg	aagatcaaac	ccatacaaac	900
ggtcatgttg	ttcatgatgc	acagcatcct	tcaagatcta	atgtggcttt	caaaaaggaa	960
ctacacacct	ctcaatcaga	taggtcttgt	actcacttag	aattcgatat	ttctcacaca	1020
ggactgtctt	acgaaactgg	cgatcacgtt	ggcgtttatt	ccgagaactt	gtccgaagtt	1080
gtcgatgaag	cactaaaaact	gttaggggta	tcaccagaca	catacttctc	agtccatgct	1140
gataaggagg	atgggacacc	tatcggtggt	gcttcaactac	caccaccttt	tcctccttgc	1200
acattgagag	acgctctaac	cagatacgca	gatgtcttat	ctcaccttaa	aaaggtagct	1260
ttgtctggcat	tggctgtctca	tgctagtgtat	cctagtgaag	ccgatagggt	aaagtctctg	1320
gcttcaccag	ccggaagaa	tgaatatgca	caatggatcg	tcgccaacca	acgttctttg	1380
ctagaagtga	tgcagagttt	tccatctgcc	aagcctccat	taggtgtggt	cttcgcagca	1440
gtagctccac	gtttacaacc	aagatactac	tctatcagtt	catctcctaa	gatgtctcct	1500
aacagaatac	atgtttacatg	tgctttgggt	tacgagacta	ctccagcagg	cagaattcac	1560
agaggattgt	gttcaacctg	gatgaaaaat	gctgtccctt	taacagagtc	acctgattgc	1620
tctcaagcat	ccattttcgt	tagaacatca	aatttcagac	ttccagtggg	tccaaaagtt	1680
ccagtcatta	tgataggacc	aggcactggg	cttgcctccat	tcaggggctt	tcttcaagag	1740
agattggcct	tgaaggaact	tggtacagaa	ttgggttctt	ctatcttttt	ctttgggttg	1800
cgtaaatagaa	aagttgactt	tatctacgag	gacgagctta	acaattttgt	tgagacagga	1860
gcattgtcag	aattgatcgt	cgcattttca	agagaaggga	ctgccaaaaga	gtacgttcag	1920
cacaagatga	gtcaaaaaagc	ctccgatata	tggaaacttc	taagtgaagg	tgcttatctt	1980
tatgtctgtg	gcgatgcaaa	gggatggccc	aaggatgtcc	atagaactct	gcatacaatt	2040
gttcagggaac	aagggtgctc	gattctcttc	aaggctgaat	tgtagctcaa	aaacttacag	2100
atgtctggaa	gatacttaag	agatgtttgg	taa			2133

## SEQ ID NO:62

atgaaggcca	gtccattcga	attcatgtcc	gctattatca	agggtagaat	ggacccatct	60
aactcctcat	ttgaatctac	tggtgaagtt	gcctccgtta	tctttgaaaa	cagagaattg	120
gttgccatct	tgaccacttc	tattgctgtt	atgattgggt	gcttcgttgt	cttgatgtgg	180
agaagagctg	gttctagaaa	ggttaagaat	gtcgaattgc	caaagccatt	gattgtccat	240
gaaccagaac	ctgaagtgtg	agatggtaag	aagaagggtt	ccatcttctt	cggtaactca	300
actgggtactg	ctgaaggttt	tgctaaggct	ttggctgatg	aagctaaaagc	tagatacgaa	360
aaggctacct	tcagagttgt	tgatttggat	gattatgtcg	ccgatgatga	ccaatacgaa	420
gaaaaattga	agaacgaatc	cttcgccgtt	ttcttgggtg	ctacttatgg	tgatggtgaa	480
cctactgata	atgctgctag	attttacaag	tggttcgccc	aaggtaaaaga	aagaggtgaa	540
tggttgcaaa	acttgcacta	tgctgttttt	ggtttgggta	acagacaata	cgaacacttc	600

aacaagattg	ctaaggttgc	cgacgaatta	ttggaagctc	aaggtggtaa	tagattgggt	660
aaggttgggt	taggtgatga	cgatcaatgc	atcgaagatg	atctttctgc	ttggagagaa	720
tctttgtggc	cagaattgga	tatgttgttg	agagatgaag	atgatgctac	tactgttact	780
actccatata	ctgctgctgt	cttgggaatac	agagttgtct	ttcatgattc	tgctgatgtt	840
gctgtcgaag	ataagtcttg	gattaacgct	aatggctcatg	ctgttcatga	tgctcaacat	900
ccattcagat	ctaacgttgt	cgtcagaaaa	gaattgcata	cttctgcctc	tgatagatcc	960
tgttctcatt	tggaattcaa	catttcgggt	tccgctttga	attacgaaac	tggtgatcat	1020
gttggtgtct	actgtgaaaa	cttgactgaa	actgttgatg	aagccttgaa	cttgttgggt	1080
ttgtctccag	aaacttactt	ctctatctac	accgataacg	aagatggtac	tccattgggt	1140
ggttcttcat	tgccaccacc	atctccatca	tgtactttga	gaactgcttt	gaccagatac	1200
gctgatttgt	tgaactctcc	aaaaaagtct	gctttgttgg	ctttagctgc	tcatgcttct	1260
aatccagttg	aagctgatag	attgagatac	ttggcttctc	cagctggtaa	agatgaatat	1320
gccaatcttg	ttatcggttc	ccaaaagtct	ttgttggaag	ttatggctga	attcccatct	1380
gctaaaccac	cattaggtgt	tttttttgct	gctgttgctc	caagattgca	acctagattc	1440
tactccatct	catctctctc	aagaatggct	ccatctagaa	tccatgttac	ttgtgctttg	1500
gtttacgata	agatgccaac	tggtagaatt	cataaggggtg	tttgttctac	ctggatgaag	1560
aattctgttg	caatggaaaa	gtcccatgaa	tgttcttggg	ctccaatttt	cgttagacaa	1620
tccaatttta	agttgcagc	cgaaatccaag	gttccaatta	tcatggttgg	tccaggtaact	1680
ggtttggctc	cttttagagg	ttttttacaa	gaaagattgg	ccttgaaaga	atccgggtgtt	1740
gaattgggtc	catccatttt	gtttttcggt	tgcaagaaaca	gaagaatgga	ttacatctac	1800
gaagatgaat	tgaacaactt	cgttgaaacc	gggtgctttgt	ccgaattggt	tattgctttt	1860
tctagagaag	gtccctaccac	agaatacgtc	caacataaga	tggctgaaaa	ggcttctgat	1920
atctggaact	tgatttctga	agggtgcttac	ttgtacgttt	gtggtgatgc	taaaggtatg	1980
gctaaggatg	ttcatagaac	cttgcatacc	atcatgcaag	aacaaggttc	tttggtattct	2040
tccaaagctg	aatccatggt	caagaacttg	caaatgaatg	gtagatactt	aagagatggt	2100
tggttaa						2160

## SEQ ID NO:63

aagcttaaaa	tgagtaagtc	taatagtatg	aattctacat	cacacgaaac	cctttttcaa	60
caatttgggt	tggttgggga	ccgtatgcc	ttgatggatg	ttcactgggt	gatctacgtt	120
gctttcggcg	catggttatg	ttcttatgtg	atacatgttt	tatcatcttc	ctctacagta	180
aaagtgcag	ttgttggata	caggctctgta	ttcgaacct	catggttgct	tagacttaga	240
ttcgtctggg	aaggtggctc	tatcataggt	caagggtaca	ataagtttaa	agactctatt	300
ttccaagtta	ggaaattggg	aactgatatt	gtcattatac	cacctaaact	tattgatgaa	360
gtgagaaaa	tgtcacagga	caagactaga	tcagtgtgaa	ctttcattaa	tgattttgca	420
ggtcaatata	caagaggcat	ggttttcttg	caatctgact	tacaaaaccg	tggtatacaa	480
caaagactaa	ctccaaaatt	ggtttccctg	accaagggtc	tgaagggaag	gttggtattt	540
gctttaacaa	aagagatgcc	tgatatgaaa	aatgacgaat	gggtagaagt	agatatcagt	600
agtataatgg	tgagattgat	ttccaggatc	tccgccagag	tctttctagg	gcctgaacac	660
tgctgtaacc	aggatgggt	gactactaca	gcagaatatt	cagaatcact	tttcattaca	720
gggtttatct	taagagttgt	acctcatatc	ttaagacct	tcatcgcccc	tctattacct	780
tcatacagga	ctctacttag	aaacgtttca	agtggtagaa	gagtcacgg	tgacatcata	840
agatctcagc	aaggggatgg	taacgaagat	atactttcct	ggatgagaga	tgctgccaca	900
ggagagggaa	agcaaatcga	taacattgct	cagagaatgt	taattctttc	tttagcatca	960
atccacacta	ctgcatgac	catgacacat	gccatgtacg	atctatgtgc	ttgccctgag	1020
tacattgaac	cattaagaga	tgaagttaaa	tctgttgggt	gggcttctgg	ctgggacaag	1080
acagcggtta	acagatttca	taagttggac	tccttcctaa	aagagtcaca	aagattcaac	1140
ccagtattct	tattgacatt	caatagaatc	taccatcaat	ctatgacctt	atcagatggc	1200
actaacattc	catctggaac	acgtattgct	gttccatcac	acgcaatgtt	gcaagattct	1260
gcacatgtgc	caggtccaac	ccacacctact	gaatttgatg	gattcagata	tagtaagata	1320
cgttctgata	gtaactcagc	ccaaaagtac	ctattctcca	tgaccgatcc	ttcaaacatg	1380
gctttcggat	acggcaagta	tgcttgctca	ggtagatttt	acgcgtctaa	tgagatgaaa	1440
ctaacattag	ccattttgtt	gctacaattt	gagttcaca	taccagatgg	taaaggctcg	1500
cctagaaaata	tactatcga	ttctgatatg	attccagacc	caagagctag	actttgcgtc	1560
agaaaaagat	cacttagaga	tgaatgaccg	cgg			1593

## SEQ ID NO:64

aagcttaaaa	tggaagatcc	tactgtctta	tatgcttgct	ttgccattgc	agttgcaact	60
ttcgttggta	gatggtacag	agatccattg	agatccatcc	caacagtttg	tggttccgat	120
ttgcctattc	tatcttacat	cggcgacta	agatggacaa	gacgtggcag	agagatactt	180
caagagggat	atgatggcta	cagaggatct	acattcaaaa	tcgcatgtgt	agaccgtttg	240
atcgtgatcg	caaatgggtc	taaactagct	gatgaagtca	gacgtagacc	agatgaagag	300
ttaaacttta	tgagcggatt	aggagcattc	gtccaaacta	agtagacact	aggtgaagct	360

attcataacg	atccataacca	tgtcgatatac	ataagagaaa	aactaacaag	aggccttcca	420
gccgtgcttc	ctgatgtcat	tgaagagtgg	acacttgccg	ttagacagta	cattccaaca	480
gaaggtgatg	aatgggtgtc	cgtaaaactgt	tcaaaggccg	caagagatat	tgttgctaga	540
gcttctaata	gagtcctttgt	aggtttgccct	gcttgcgaaa	accaagggtta	cttagatttg	600
gcaatagact	ttacattgtc	tgttgtcaag	gatagagcca	tcatcaatat	gtttccagaa	660
ttgttgaagc	caatagttgg	cagagtgtga	ggtaacgcca	ccagaaatgt	tcgtagagct	720
gttccttttg	ttgctccatt	gggtggaggaa	agacgtagac	ttatggaaga	gtacggtgaa	780
gactggctcg	aaaaacctaa	tgatatgtta	cagtggtata	tggtatgaagc	tgcatccaga	840
gatagttcag	tgaaggcaat	cgcagagaga	ttgttaatgg	tgaacttcgc	ggctattcat	900
acctcatcaa	acactatcac	tcattgctttg	taccaccttg	ccgaaatgcc	tgaaactttg	960
caaccactta	gagaagagat	cgaaccatta	gtcaaagagg	agggctggac	caaggctgct	1020
atgggaaaaa	tgtgtgtggt	agattcattt	ctaagagaat	ctcaaagata	caatggcatt	1080
aacatcgat	ctttaactag	aatggctgac	aaagatatta	catttgagtga	tggcacattt	1140
ttgccaaaa	gtactctagt	ggccgttcca	gcgtattcta	ctcatagaga	tgatgctgtc	1200
cagctgatg	ccttagtatt	cgatcctttc	agattctcac	gtatgagagc	gagagaaggt	1260
gaaggtacaa	agcaccagtt	cgtaataact	tcagtcgagt	acgttccatt	tggtcacgga	1320
aagcatgctt	gtccaggaa	attcttcgcc	gcaaacgaat	tgaaagcaat	gttggcttac	1380
attgttctaa	actatgatgt	aaagttgcct	ggtagcggta	aacgtccatt	gaacatgtat	1440
tggggtccaa	cagttttgcc	tgcaaccagca	ggccaagtat	tgttcagaaa	gagacaagtt	1500
agtcataaac	cgcg					1515

## SEQ ID NO:65

atggctacct	tgttgaaca	ttttcaagct	atgccattcg	ctattccaat	tgctttggct	60
gctttgtctt	gggtgtttt	gttctacatc	aaggtttctt	tcttctccaa	caaatccgct	120
caagctaaat	tgccaccagt	tccagttgtt	ccaggtttgc	cagttattgg	taatttgttg	180
caattgaaag	aaaagaagcc	ataccaaacc	ttcactagat	gggctgaaga	atatggtcca	240
atctactcta	ttagaactgg	tgcttctact	atggttgtct	tgaacactac	tcaagttgcc	300
aaagaagcta	tggttaccag	atacttgtct	atctctacca	gaaagtgtgc	caacgccttg	360
aaaaatttga	ccgctgataa	gtgcatgggt	gccatttctg	attacaacga	tttccacaag	420
atgatcaaga	gatatactct	gtctaacggt	ttgggtccat	ctgccccaaa	aagacataga	480
tctaacagag	ataccttgag	agccaacggt	tgttctagat	tgcatcccca	agttaagaac	540
tctccaagag	aagctgtcaa	ctttagaaga	gttttcgaat	gggaattatt	cggtatcgct	600
ttgaaacaag	ccttcggtaa	ggatattgaa	aagccaatct	acgtcgaaga	atttgggtact	660
actttgtcca	gagatgaaat	cttcaaggtt	ttggtcttgg	acattatgga	aggtgccatt	720
gaagttgatt	ggagagattt	ttcccatatc	ttgcgttgg	ttccaaacac	cagaatggaa	780
actaagatcc	aaagattata	ctttagaaga	aaggccgtta	tgaccgcctt	gattaacgaa	840
caaaagaaaa	gaattgcctc	cggtagaaga	atcaactgct	acatcgattt	cttgttgaaa	900
gaaggttaaga	ccttgaccat	ggaccaaata	tctatgttgt	tgtgggaaac	cgttattgaa	960
actgctgata	ccacaatggt	tactactgaa	tgggctatgt	acgaagtgtg	taaggattct	1020
aaaagacaag	acagattata	ccaagaaatc	caaaaggctc	gcggttctga	aatggttaca	1080
gaagaatact	tgtcccaatt	gccatacttg	aatgctgttt	tccacgaaac	tttgagaaaa	1140
cattctccag	ctgcttttgt	tccattgaga	tatgctcatg	aagatactca	atttgggtgt	1200
tattacattc	cagccggtac	tgaatttgcc	attaacatct	acggttgcaa	catggacaaa	1260
caccaatggg	aattctccga	caaatggaag	ccagaaagat	ttttggatcc	taagtttgac	1320
ccaatggact	tgtacaaaac	tatggctttt	gggtctggta	aaagagtttg	cgctggttct	1380
ttacaagcta	tgttgattgc	ttgtccaacc	atcggtagat	tgtttcaaga	atttgaatgg	1440
aagttgagag	atgggtgaaga	agaaaacggt	gatactgttg	gtttgaccac	ccataagaga	1500
tatccaatgc	atgctatttt	gaagccaaga	tcttaa			1536

## SEQ ID NO:66

atggcagaat	tagatacact	tgatatagta	gtattaggtg	ttatcttttt	gggtactgtg	60
gcatacttta	ctaagggttaa	attgtggggt	gttaccaagg	atccatacgc	taacggattc	120
gctgcagggtg	gtgcttccaa	gcctggcaga	actagaaaaca	tcgtcgaagc	tatggaggaa	180
tcaggtaaaa	actgtgttgt	ttctacggc	agtcaaaacag	gtacagcgga	ggattacgca	240
tcaagacttg	caaagggaag	aaagtccaga	ttcggtttga	acactatgat	cgccgatcta	300
gaagattatg	acttcgataa	cttagacact	gttccatctg	ataacatcgt	tatgtttgta	360
ttggctactt	acgggtgaag	cgaaccaaca	gataacgccg	tggtatttcta	tgagttcatt	420
actggcgaag	atgcctcttt	caatgagggc	aacgatcctc	cactaggtaa	cttgaattac	480
gttgcggttcg	gtctgggcaa	caatacctac	gaacactaca	actcaatggt	caggaacggt	540
aacaaggctc	tagaaaaagt	aggagctcat	agaattggag	aagcaggtga	gggtgacgac	600
ggagctggaa	ctatggaaga	ggacttttta	gcttggaag	atccaatgtg	ggaagccttg	660
gctaaaaaga	tgggcttgga	ggaaagagaa	gctgtatatg	aacctatttt	cgctatcaat	720
gagagagatg	atttgacccc	tgaagcgaat	gaggtatact	tgggagaacc	taataagcta	780

cacttggaag	gtacagcgaa	aggctccatc	aactcccaca	acccatatat	cgcaccaatt	840
gcagaatcat	acgaactttt	ctcagctaag	gatagaaatt	gtctgcata	ggaaattgat	900
atttctggta	gtaatctaaa	gtatgaaaca	ggcgaccata	tcgcgatctg	gcctaccaac	960
ccaggtgaag	aggtaacaaa	atttcttgac	attctagatc	tgtctggtaa	gcaacattcc	1020
gtcgtaacag	tgaagcctt	agaacctaca	gccaaagttc	cttttccaaa	tccaactacc	1080
tacgatgcta	tattgagata	ccatctggaa	atatgcgctc	cagtttctag	acagtttgtc	1140
tcaactttag	cagcattcgc	ccctaattgat	gatatcaaa	ctgagatgaa	ccgtttggga	1200
tcagacaaag	attacttcca	cgaaaagaca	ggaccacatt	actacaatat	cgctagattt	1260
ttggcctcag	tctctaagg	tgaaaaatgg	acaaagatac	cattttctgc	tttcatagaa	1320
ggccttacaa	aactacaacc	aagatactat	tctatctctt	cctctagttt	agttcagcct	1380
aaaaagatta	tgattactgc	tggtgtcgaa	tctcagcaaa	ttccaggtag	agatgacca	1440
ttcagaggtg	tagcgactaa	ctacttggtc	gctttgaagc	agaaacaaaa	cggtgatcca	1500
aatccagctc	cttttgccca	atcatcacgag	ttgacaggac	caagggaataa	gtatgatggt	1560
atacatgttc	cagtcctatg	aagacattct	aactttaagc	taccatctga	tccaggcaaa	1620
octattatca	tgatcgggtc	aggtaaccgt	gttgcccttt	ttagaggctt	cgccaagag	1680
agggcaaaac	aagccagaga	tggtgtagaa	gttggttaaaa	cactgctggt	ctttggatgt	1740
agaaagagta	cagaagattt	catgtatcaa	aaagagtggc	aagagtacaa	ggaagctctt	1800
ggcgacaaat	tcgaaatgat	tacagctttt	tcaagagaag	gatctaaaaa	ggtttatggt	1860
caaacagattg	ctaaagaaag	atcaaaaggaa	gtttctgatac	ttctatccca	aaaagcatac	1920
ttctacgttt	gcggagacgc	gcacacatg	gcacgtgaag	tgaacactgt	gttagcacag	1980
atcatagcag	aaggccgtgg	tgtatcagaa	gccaaagggtg	aggaaattgt	caaaaacatg	2040
agatcagcaa	atcaatacca	agtgtgttct	gatttcgtaa	ctttacactg	taaagagaca	2100
acatacgcca	attcagaatt	gcaagaggat	gtctggagtt	aa		2142

## SEQ ID NO:67

atggccgaat	tggatacctt	ggatatcggt	gttttggtg	ttatcttctt	gggtactggt	60
gcttacttca	ccaaaggtaa	attgtggggt	gttactaagg	atccatacgc	taatggtttt	120
gctgtcgggt	gtgcttctaa	accaggtaga	actagaaata	tcggtgaagc	catggaagaa	180
tctggtaaga	actgtgttgt	tttctacggt	tctcaaaactg	gtactgctga	agattatgct	240
tccagatttg	ctaaagaagg	taagagtaga	ttcggtttga	acaccatgat	tgccgatttg	300
gaagattacg	atctcgataa	cttgataacc	gtcccactctg	ataacatcgt	tatgtttggt	360
ttggctacct	acgggtgaag	tgaacctact	gataatgctg	ttgacttcta	cgaattcatt	420
accggtgaag	atgcttcttt	caacgaaggt	aatgatccac	cattgggtaa	cttgaattac	480
gttgcttttg	gtttgggtaa	caacacctac	gaacattaca	actccatggt	tagaaacgtc	540
aacaaggctt	tggaataatt	gggtgctcat	agaattgggtg	aagctgggtg	agggtgatgt	600
ggtgtcggta	ctatggaaga	agattttttg	gcttggaag	acccaatgtg	ggaagccttg	660
gctaaaaaga	tgggttttga	agaaagagaa	gctgtctacg	aacctatttt	cgccattaac	720
gaaagagatg	atttgacccc	tgaagccaat	gaagtttatt	tgggtgaacc	taacaagttg	780
cacttggaag	gtactgctaa	aggctcattc	aattctcaca	acccatatat	tgctccaatc	840
gccgaatctt	acgaattatt	ctctgctaag	gatagaaact	gcttgcacat	ggaaattgac	900
atctctggtt	ctaatttgaa	gtacgaacc	ggtgatcata	ttgccatttg	gccaactaat	960
ccaggtgaag	aagttaacaa	gttcttggtc	atcttggtg	tgtccggtaa	acaacattct	1020
gttggttactg	ttaaggcctt	ggaacctaca	gctaaagttc	cttttccaaa	tccaactacc	1080
tacgatgccca	ttttgagata	ccatttgtaa	atttgcgctc	cagtctctag	acaattcggt	1140
tctacttttg	ctgcttttgc	tccaaaacgat	gatattaagg	ctgaaatgaa	cagattgggt	1200
tccgataaag	attacttcca	cgaaaaaact	ggtccacact	actacaacat	tgctagattt	1260
ttggcctctg	tctctaaagg	tgaaaagtgg	actaaagattc	cattctccgc	tttcattgaa	1320
ggtttgacta	agttgcaacc	tagatattac	tccatctcct	cctcatcttt	ggttcaacct	1380
aagaagatct	ctattaccgc	cgttggtgaa	tcccaacaaa	ttccaggtag	agatgatcct	1440
tttagagggtg	ttgctaccaa	ttacttggtc	gccttgaaac	aaaagcaaaa	cggtgatcca	1500
aatcctgctc	catttggtca	atcttatgaa	ttgactgggtc	caagaaacaa	gtacgatggt	1560
atcctatgttc	cagttcacgt	tagacactct	aactttaagt	tgccatctga	tccaggtaag	1620
ccaattatca	tgattggtcc	aggtactggt	gttgctccat	tcagagggtt	tggtcaagaa	1680
agagctaagc	aagctagaga	tggtgttgaa	gttggttaaaa	ccttggtggt	cttcggttgt	1740
agaaagtcca	ctgaagattt	catgtaccaa	aaagaatggc	aagaatacaa	agaagcctta	1800
ggtgacaagt	tcgaaatgat	tactgccttc	tcaagagaag	gttctaagaa	ggtttacgtc	1860
caacacagat	tgaaagaaag	atccaaagaa	gtctccgatt	tggtgtotca	aaaggcctac	1920
ttttacgttt	gtggtgatgc	tgctcatatg	gccagagaag	ttaatactgt	tttggtccaa	1980
attatcgctg	aaggtagagg	tgtatctgaa	gctaagggtg	aagaaatcgt	taagaacatg	2040
agatccgcca	atcaatacca	agtttgctct	gattttgtta	ccttgcaactg	taaagaaacc	2100
acctacgcta	attccgaatt	gcaagaagat	gtttggtcct	aa		2142

## SEQ ID NO:68



MEASYLYISI	LLLLASYLFT	TQLRRKSANL	PPTVFPSIPI	IGHLYLLKKP	LYRTLAKIAA	60
KYGPILQLQL	GYRRVLVISS	PSAAEECFN	NDVIFANRPK	TLFGKIVGGT	SLGSLSYGDO	120
WRNLRRVASI	EILSVHRLNE	FHDIRVDENR	LLIRKLRSST	SPVTLITVFI	ALTINVIMRM	180
ISGKRYFDSG	DRELEEEGKR	FREILDETL	LAGASNVGDY	LPILNWLGVK	SLEKKLIALQ	240
KKRDDFFQGL	IEQVRKSRGA	KVGKGRKMTI	ELLLSLQESE	PEYYTDAMIR	SFVLGLLAAG	300
SDTSAGTMEW	AMSLLVNHPH	VLKKAQAEID	RVIGNNRLID	ESDIGNIPYI	GCIINETLRL	360
YPAGPLLFPH	ESSADCVISG	YNIPRGTMIL	VNQWAIHHPD	KVWDDPETFK	PERFQGLEGT	420
RDGFKLMPFG	SGRRGCPGEG	LAIRLLGMTL	GSVIQCDFWE	RVGDEMVDMT	EGLGVTLPKA	480
VPLVAKCKPR	SEMTNLLSEL					500

## SEQ ID NO:69

MQSESVEAST	IDLMTAVLKD	TVIDTANASD	NGDSKMPPAL	AMMFEIRDLL	LILTTSAVL	60
VGCFVVLVWK	RSSGKKSGKE	LEPPKIVVPK	RRLEQEVDDG	KKKVTIFFGT	QTGTAEFPAK	120
ALFEEAKARY	EKAAFKVIDL	DDYAADLDEY	AEKLLKETYA	FFFLATYGDG	EPTDNAKFY	180
KWFTGDEKQ	VWLQKLQYGV	FGLGNRQYEH	FNKIGIVVDD	GLTEQGAARI	VPVGLGDDQ	240
SIEDDFSAAK	ELVWPELDLL	LRDEDDKAAA	TPYTAAIPEY	RVVFHDKPDA	FSDHTQTNG	300
HAVHDAQHPC	RSNVAVKKEL	HTPESDRSCT	HLEFDISHTG	LSYETGDHVG	VYCENLIEVV	360
EEAGKLLGLS	TDTYFSLHID	NEDGSPLGGP	SLQPPFPCT	LRKALTNYAD	LLSSPKKSTL	420
LALAAHASDP	TEADRLRFLA	SREGKDEYAE	WVVANQRSLL	EVMEAFPSAR	PPLGVFFAAV	480
APRLQPRYYS	ISSSPKMEPN	RIHVTCALVY	EKTPAGRIHK	GICSTWMKNA	VPLTESQDCS	540
WAPIFVRTSN	FRLPIDPKVP	VIMIGPGTGL	APFRGFLQER	LALKESGTEL	GSSILFFGCR	600
NRKVDYIYEN	ELNNEFVNGA	LSELDVAFSR	DGPTKEYVQH	KMTQKASEIW	NMLSEGAYLY	660
VCDAKGMMAK	DVHRTLHTIV	QEQGSLDSSK	AELYVKNLQM	SGRYLRDVW		709

## SEQ ID NO:70

MASITHFLQD	FQATPFATAF	AVGVSVLLIF	FFFIRGFHST	KKNEYKLP	VPVVPGLPVV	60
GNLLQLKEKK	PYKTLRWAE	IHGPIYSIRT	GASTMVVNS	THVAKEAMVT	RFSSISTRKL	120
SKALELLTSN	KSMVATSDYN	EFHKMVKKYI	LAELLGANAQ	KRHRIRHRTL	IENVLNKLHA	180
HTKNSPLQAV	NFRKIFESL	FGLAMKQALG	YDVSLEFVEE	LGTTLSREEI	YNVLVSDMLK	240
GAEVVDWRDF	FPYLKWIPNK	SFEMKIQLRA	SRQAVMNSI	VKEQKKSIA	GKGENCYLNY	300
LLSEAKTLTE	KQISILAWET	IIETADTTTV	TTEWAMYELA	KNPKQDRLY	NEIQNVCGTD	360
KITEEHL SKL	PYLSAVFHET	LRKYSFSPVL	PLRYAHEDTQ	LGGYVVPAGT	EIAVNIYGCN	420
MDKNQWETPE	EWKPERFLDE	KYDFMDMYKT	MSFGSGKRV	AGSLQASLIA	CTSIGRLVQE	480
FEWRLKDG EV	ENVDTLGLTT	HKLYPMQAIL	QPRN			514

## SEQ ID NO:71

MASISLLLG	FVSSFLFIF	FLKKLLFFFS	RHKMSEVSRL	PSVPVPGFPL	IGNLLQLKEK	60
KPHKTFTKWS	ELYGPIYSIK	MGSSSLIVLN	SIETAKEAMV	SRFSSISTRK	LSNALTVLTC	120
NKSMVATSDY	DDFHKFVKRC	LLNGLLGANA	QERKRHYRDA	LIENVTSKLH	AHTRNHPQEP	180
VNFRAIFEHE	LFGVALKQAF	GKDVESIYVK	ELGVTLSRDE	IFKVLVHDM	EGAIQVVDWRD	240
FFPYLKWIPN	NSFEARIQKQ	HKRLAVMNA	LIQDRNLQND	SESDDDCYLN	FLMSEAKTLT	300
MEQIAILVWE	TIETADTTTL	VTEWAMYEL	AKHQSVQDRL	FKEIQSVCGG	EKIKEEQQLPR	360
LPYVNGVFHE	TLRKYSAPPL	VPIRYAHEDT	QIGGYHIPAG	SEIAINIYGC	NMDKKRWERP	420
EEWWPERFLE	DRYESSDLHK	TMAFGAGKRV	CAGALQASLM	AGIAIGRLVQ	EFEWKL RDGE	480
EENVDTYGLT	SQKLYPLMAI	INPRRS				506

## SEQ ID NO:72

MDMMGIEAVP	FATAVVLGGI	SLVVLIFIRR	FVSNRKRVS	GLPPVPDIPG	LPLIGNLLQL	60
KEKKPHKTFA	RWAETYGPIF	SIRTGASTMI	VLNSSEVAKE	AMVTRFSSIS	TRKLSNALKI	120
LTFDKCMVAT	SDYNDFHMKV	KGFILRNVLG	APAQKRHRCH	RDTLIENISK	YLHAHVKTSP	180
LEPVVLKKIF	ESEIFGLALK	QALGKDIESI	YVEELGTTLS	REEIFAVLVV	DPMAGAEVD	240
WRDFFPYLSW	IPNKSMEKMI	QRMDFRRGAL	MKALIGEQQK	RIGSGEEKNS	YIDFLLSEAT	300
TLTEKQIAML	IWETIIIEISD	TTLVTSEWAM	YELAKDPNRQ	EILYREIHKV	CGSNKLTEEN	360
LSKLPYLSNV	FHETLRKYSP	APMVPVRYAH	EDTQLGGYHI	PAGSQIAINI	YGCNMNKKQW	420
ENPEEWKPER	FLDEKYDLM	LHKTMAFGGG	KRVACALQA	MLIACTSIGR	FVQEFEWKLM	480
GGEEENVDTV	ALTSQKLHPM	QAIKARE				508

## SEQ ID NO:73

MAELDTLDIV	VLGVIFLGTV	AYFTKGKLG	VTKDPYANGF	AAGGASKPGR	TRNIVEAMEE	60
SGKNCVVFY	SQTGTAEDYA	SRLAKEGKSR	FGLNTMIADL	EDYDFDNLDT	VPNDNIVMFV	120
LATYGEGET	DNAVDFYEFI	TGEDASFNEG	NDPPLGNLNY	VAFLGLNNTY	EHYNSMVRNV	180
NKALEKLGAH	RIGEAGEGDD	GAGTMEEDFL	AWKDPMWEAL	AKKMGLEERE	AVYEPIFAIN	240
ERDDLTPEAN	EVYLGEPNKL	HLEGTAAGPF	NSHNPYIAP	AESYELFSAK	DRNCLHMEID	300
ISGSNLKYET	GDHIAIWPTN	PGEEVNKFLD	ILDLSGKQHS	VVTVKALEPT	AKVFPNPPTT	360
YDAILRYHLE	ICAPVSRQFV	STLAFAFAPND	DIKAEMNRLG	SDKDYFHEKT	GPHYNYIARF	420
LASVSKGEKW	TKIPFSAFIE	GLTKLQPRYY	SISSSSLVQP	KKISITAVVE	SQIIPGRDDP	480
FRGVATNYLF	ALKQKQNGDP	NPAPFGQSYE	LTGPRNKYDG	IHPVHVHRHS	NFKLPSDPGK	540
PIIMIGPGTG	VAPFRGFVQE	RAKQARDGVE	VGKTLFFGFC	RKSTEDFMYQ	KEWQEYKEAL	600
GDKFEMITAF	SREGSKKVYV	QHRLKERSKE	VSDLLSQKAY	FYVCGDAAHM	AREVNTVLAQ	660
IIAEGRGVSE	AKGEEIVKNM	RSANQYQVCS	DFVTLLHCKET	TYANSELQED	VWS	713

## SEQ ID NO:74

MKVSPFEFMS	AIKGRMDPS	NSSFESTGEV	ASVIFENREL	VAILTTSIAV	MIGCFVVLW	60
RRAGSRKVKN	VELPKPLIVH	EPEPEVEDGK	KKVSIFFGTQ	TGTAEGFAKA	LADAEKARYE	120
KATFRVVDLD	DYAADDQYE	EKLKNESFAV	FLLATYGDGE	PTDNAARFYK	WFAEGKERGE	180
WLQNLHYAVF	GLGNRQYEHF	NKIAKVADEL	LEAQGGNRLV	KVGLGDDDDQ	IEDDFSARWE	240
SLWPELMDLL	RDEDDATTVT	TPYTAADVLEY	RVVFHDSADV	AAEDKSWINA	NGHAVHDAQH	300
PPRSNVVVRK	ELHTSASDRS	CSHLEFNISG	SALNYETGDH	VGVCENLITE	TVDEALNLLG	360
LSPETYFSIY	TDNEDGTPLG	GSSSLPPFPFS	CTLRTALTRY	ADLLNSPKKS	ALLALAAHAS	420
NPVEADRLRY	LASPAKGDEY	AQSVIGSQKS	LLEVMAEFPS	AKPPLGVFFA	AVAPRLQPRF	480
YSISSSPRMA	PSRIHVTCAL	VYDKMPTGRI	HKGVCSTWMK	NSVPMESKSE	CSWAPIFVRQ	540
SNFKLPAESK	VPIIMVGPPT	GLAPFRGFLQ	ERLALKESGV	ELGPSILFFG	CRNRRMDYIY	600
EDELNNFVET	GALSELVIAF	SREGPTKEYV	QHKMAEKASD	IWNLISEGAY	LYVCGDAKGM	660
AKDVHRTLHT	IMQEQGSLDS	SKAESMVKNL	QMNGRYLRDV	W		701

## SEQ ID NO:75

MATLLEHFQA	MPFAIPIALA	ALSWLFLFYI	KVSFFSNKSA	QAKLPPVPVV	PGLPVIGNLL	60
QLKEKKPYQT	FTRWAEYGP	IYSIRTGAST	MVVLNTTQVA	KEAMVTRYLS	ISTRKLSNAL	120
KILTADKCMV	AISDYNDHFH	MIKRYILSNV	LGPSAQKRHR	SNRDLTRANV	CSRLHSQVKN	180
SPREAVNFR	VFEWELFGIA	LKQAFGKDIE	KPIYVEELGT	TLRDEIFKV	LVLDIMEGAI	240
EVDWRDFPHY	LRWIPNTRME	TKIQRLYFRR	KAVMTALINE	QKKRIASGEE	INCYIDFLK	300
EGKTLTMDQI	SMLLWETVIE	TADTTMTVTE	WAMYEVAKDS	KRQDRLYQEI	QKVCSEMVT	360
EEYLSQLPYL	NAVHETLRK	HSPAALVPLR	YAHEDTQLGG	YYIPAGTEIA	INIYGCNMDK	420
HQWESPPEWK	PERFLDPKFD	PMDLYKTMAF	GAGKRVCAQS	LQAMLIACPT	IGRLVQEFEE	480
KLRDGEENNV	DTVGLTTHKR	YPMHAILKPR	S			511

## SEQ ID NO:76

MQSDSVKVSP	FDLVSAMNG	KAMEKLNASE	SEDPTTLPAL	KMLVENRELL	TLFTTSFAVL	60
IGCLVFLMWR	RSSSKKLVD	PVPQVIVVKK	KEKESEVDDG	KKKVSIFYGT	QTGTAEGFAK	120
ALVEEAKVRY	EKTSFKVIDL	DDYAADDDEY	EEKLKESLA	FFFLATYGDG	EPTDNAANFY	180
KWFTGDDKG	EWLKKLQYGV	FGLGNRQYEH	FNKIAIVVDD	KLTEMGAKRL	VPVGLGDDDDQ	240
CIEDDFTAWK	ELVWPELDQL	LRDEDDTSVT	TPYTAADVLEY	RVVYHDKPAD	SYAEDQHTN	300
GHVVHDAQHP	SRSNVAFKKE	LHTSQSDRSC	THLEFDISHT	GLSYETGDHV	GVYSENLSV	360
VDEALKLLGL	SPDTYFSVHA	DKEDGTPIGG	ASLPPFPFPC	TLRDALTRYA	DVLSPPKKVA	420
LLALAAHASD	PSEADRLKFL	ASPAGKDEYA	QWIVANQSL	LEVMSQSFPSA	KPPLGVFFAA	480
VAPRLQPRYY	SISSSPKMS	NRIHVTCALV	YETTPAGRIH	RGLCSTWMKN	AVPLTESPDC	540
SQASIFVVRTS	NFRLPVDPKV	PVIMIGPGTG	LAPFRGFLQE	RLALKESGTE	LGSSIFFFGC	600
RNRKVDFFIY	DELNNFVETG	ALSELIVAFS	REGTAKEYVQ	HKMSQKASDI	WKLLSEGAYL	660
YVCGDAKGMA	KDVHRTLHTI	VQEQGSLDSS	KAELYVKNLQ	MSGRYLRDVW		710

## SEQ ID NO:77

MSKSNSMNST	SHETLFQQLV	LGLDRMPLMD	VHWLIYVAFG	AWLCSYVIHV	LSSSSTVKVP	60
VVGYSRVFEP	TWLLRLRFVW	EGGSIIGQGY	NKFKDSIFQV	RKLGTDIVII	PPNYIDEVRK	120
LSQDKTRSV	PFINDFAGQY	TRGMVFLQSD	LQNRVIQORL	TPKLVSILTKV	MKEELDYALT	180
KEMPDMMKND	WVEVDISSIM	VRILSRISAR	VFLGPEHCRN	QEWLTTTAEY	SESLFITGFI	240
LRVVPILHRP	FIAPLLPSYR	TLLRNVSSGR	RVIGDIIRSQ	QGDGNEDILS	WMRDAATGEE	300
KQIDNIAQRM	LILSLASLHT	TAMTMTHAMY	DLACACEYIE	PLRDEVKSVV	GASGWDKTAL	360
NRFHKLDSEF	KESQRFNPVF	LLTFNRHYHQ	SMTLSDGNTI	PSGTIRIAPVS	HAMLQDSAHV	420

PGPTPTTEFD	GFRYSKIRSD	SNYAQKYLFS	MTDSSNMAFG	YGKYACPGRF	YASNEMKLTL	480
AILLLQFEFK	LPDGKGRPRN	ITIDSDMIPD	PRARLCVRKR	SLRDE		525

## SEQ ID NO:78

MEDPTVLYAC	LAIIVATFV	RWYRDLRSI	PTVGGSDLPI	LSYIGALRWT	RRGREILQEG	60
YDGYRGSTFK	IAMLDRIWIV	ANGPKLADEV	RRRPDEELNF	MDGLGAFVQT	KYTLGEAIHN	120
DPYHVDIIRE	KLTRGLPAVL	PDVIEELTLA	VRQYIPTEGD	EWVSVNCSKA	ARDIVARASN	180
RVFVGLPACR	NQGYLDLAI	FTLSVVKDRA	IINMFPELLK	PIVGRVVGNA	TRNVRRVVPF	240
VAPLVEERRR	LMEEYGEDWS	EKPNDMLQWI	MDEAASRDSS	VKAIAERLLM	VNFAAIHTSS	300
NTITHALYHL	AEMPETLQPL	REEIEPLVKE	EGWTKAAMGK	MWWLDSFLRE	SQRYNGINIV	360
SLTRMADKDI	TLSDGTFLPK	GTLVAVPAYS	THRDDAVYAD	ALVDFPFRFS	RMRAREGEGT	420
KHQFVNTSVE	YVPFGHGKHA	CPGRFFAANE	LKAMLAYIVL	NYDVKLPDGD	KRPLNMYWGP	480
TVLPAAGQV	LFRKRQVSL					499

## SEQ ID NO:79

MDAVTGLLT	PATAITIGGT	AVALAVALIF	WYLSYTSAR	RSQSNHLPRV	PEVPGVPLLG	60
NLLQLKEKKP	YMTTRWAAT	YGPIYSIKTG	ATSMVVSSN	EIAKEALVTR	FQSISTRNLS	120
KALKVLTADK	TMVAMSDYDD	YHKTVKRHL	TAVLGPNAQK	KHRIHRDIMM	DNISTQLHEF	180
VKNNPEQEEV	DLRKIFQSEL	FGLAMRQALG	KDVESLYVED	LKITMNRDEI	FQVLVDPMM	240
GAIDVDWRDF	FPYLKWPVK	KFENTIQQMY	IRREAVMKSL	IKHEKKRIAS	GEKLSYIDY	300
LLSEAQTITD	QQLMSLWEP	IISSDITTMV	TTEWAMYELA	KNPKLQDRLY	RDIKSVCGSE	360
KITEEHLSQL	PYITAIFHET	LRRHSPVPII	PLRHVHEDTV	LGGYHVPAGT	ELAVNIYGCN	420
MDKNVWENPE	EWNPERFMKE	NETIDFQKTM	AFGGGKRVCA	GSLQALLTAS	IGIGRMVQEF	480
EWKLKDMTQE	EVNTIGLTTQ	MLRPLRAIIK	PRI			513

## SEQ ID NO:80

atggaagtaa	cagtagctag	tagtgtagcc	ctgagcctgg	tctttattag	catagtagta	60
agatgggcat	ggagtgtggt	gaattgggtg	tggtttaagc	cgaagaagct	ggaaagattt	120
ttgagggagc	aaggccttaa	aggcaattcc	tacaggtttt	tatatggaga	catgaaggag	180
aactctatcc	tgctcaaaac	agcaagatcc	aaacccatga	acctctccac	ctcccatgac	240
atagcacctc	aagtcacccc	ttttgtcgac	caaacctgta	aagcttacgg	taagaactct	300
tttaattggg	ttggcccat	accaagggtg	aacataatga	atccagaaga	tttgaaggac	360
gtcttaacaa	aaaatgttga	ctttgttaag	ccaatatcaa	accacttat	caagttgcta	420
gctacaggta	ttgcaattca	tgaagggtgag	aaatggacta	aacacagaag	gattatcaac	480
ccaacattcc	attcggaagag	gctaaagcgt	atgttacctt	catttcacca	aagttgtaat	540
gagatggta	aggaatggga	gagcttggtg	tcaaaagagg	gttcatcatg	tgagtggat	600
gtctggcctt	ttcttgaaaa	tatgtcggca	gatgtgatct	cgagaacagc	atttggaact	660
agctacaaaa	aaggacagaa	aatctttgaa	ctcttgagag	agcaagtaat	atatgtaacg	720
aaaggctttc	aaagttttta	cattccagga	tgagggttct	tcccaactaa	gatgaacaag	780
aggatgaatg	agattaacga	agaaataaaa	ggattaatca	gggtatttat	aattgacaga	840
gagcaaatca	ttaaggcagg	tgaagaaacc	aacgatgact	tattaggtgc	acttatggag	900
tcaaaactga	aggacattcg	ggaacatggg	aaaaacaaca	aaaatgttgg	gatgagtatt	960
gaagatgtaa	ttcaggagtg	taagctgttt	tactttgtcg	ggcaagaaac	cacttcagtg	1020
ttgctggcctt	ggacaatggt	tttacttggt	caaaatcaga	actggcaaga	tcgagcaaga	1080
caagagggtt	tgcaagtctt	tggaagcagc	aagccagatt	ttgatggctt	agctcacctt	1140
aaagtcgtaa	ccatgatttt	gcttgaagtt	cttcgattat	accaccaggt	cattgaactt	1200
attcgaaacca	ttcacaaaga	aacacaaact	gggaagctct	cactaccaga	aggagttgaa	1260
gtccgcttac	caacactgct	cattcaccat	gacaaggaac	tgtaggggtga	tgatgcaaac	1320
cagttcaatc	cagagaggtt	ttcggaagga	gtttccaaag	caacaaaaga	ccgactctca	1380
ttcttccctt	tcggagccgg	tccacgcatt	tgcatgggac	agaacttttc	tatgatggaa	1440
gcaaagtggg	ccttagcatt	gatottgcaa	cacttcacct	ttgagcttct	tccatctcat	1500
gcacatgctc	cttcccatcg	tataaccctt	caaccacagt	atggtgttcg	tatcatttta	1560
catcgacgtt	ag					1572

## SEQ ID NO:81

atggaagtaa	ctgtcgcttc	ttctgtcgct	ttatccttag	tcttcatttc	cattgtcgtc	60
agatgggctt	ggctccgtgt	caactgggtt	tggttcaaac	caaagaagtt	ggaaagattc	120
ttgagagagc	aagggttgaa	gggtaattct	tatagattct	tgtagcgtga	catgaaggaa	180
aattctattt	tggtgaagca	agccagatcc	aaaccaatga	acttgtctac	ctctcatgat	240
attgtccacc	aagtctactc	attcgtcgat	caaacgttta	aagcctacgg	taagaactct	300

ttcaattggg	ttgggtccaat	tcctagagtt	aacatcatga	accagaaga	tttgaaggat	360
gtcttgacca	agaacgttga	cttcgttaag	ccaatttcca	accattgat	taaattgttg	420
gctactggta	ttgccattta	cgaagggtgaa	aagtggacta	agcatagaag	aatcatcaac	480
cctaccttcc	actctgaaag	attgaagaga	atgttaccat	ctttccatca	atcctgtaat	540
gaaatggtta	aggaatggga	atccttgggt	tctaaagaag	gttcttcttg	cgaattggat	600
gtttggccat	tcttggaata	tatgtctgct	gatgtcattt	ccagaaccgc	tttcggtacc	660
tcctacaaga	agggtcaaaa	gattttcgaa	ttgttgagag	agcaagttaa	ttacgttacc	720
aagggtttcc	aatccttcta	catcccaggt	tggagattct	tgccaaactaa	aatgaacaag	780
cgtatgaacg	agatcaacga	agaaattaaa	ggtttgatca	gaggtattat	tatcgacaga	840
gaacaaatta	ttaaagctgg	tgaagaaacc	aacgatgatt	tgttggtg	tttgatggag	900
tccaacttga	aggatattag	agaacatggt	aagaacaaca	agaatgttgg	tatgtctatt	960
gaagatgtta	ttcaagaatg	taagttattc	tacttcgctg	gtcaagagac	cacttctggt	1020
ttgttagcct	ggcatatggt	cttgtaggt	caaaaccaaa	attggcaaga	tagagctaga	1080
caagaagttt	tgaagctct	cggttcttcc	aagccagact	ttgatggtt	ggcccacttg	1140
aagggttgta	ctatgatttt	gttagaagtt	ttgagattgt	accaccagt	cattgagtta	1200
atcagaacca	ttcataaaaa	gactcaattg	ggtaaattat	ctttgccaga	aggtgttgaa	1260
gtcagattac	caaccttggt	gattcaccac	gataaggaa	tatgggtg	cgacgcta	1320
caatttaatc	cagaaagatt	ttccgaaggt	gtttccaagg	ctacaaaaa	ccgtttgtcc	1380
ttcttcccat	ttggtgctgg	tccacgtatt	tgtatcggtc	aaaacttttc	catgatggaa	1440
gccaagttgg	ctttgctttt	aatcttgcaa	cacttcactt	tcgaattgtc	tccatcccat	1500
gcccacgctc	ctttctcatg	aatcacttta	caaccacaat	acggtgtcag	aatcatctta	1560
cacagaagat	aa					1572

## SEQ ID NO:82

MEVTVASSVA	LSLVFISIVV	RWAWSVVNVV	WFKPKKLERF	LREQGLKGN	YRFLYGDMKE	60
NSILLKQARS	KPMNLSTSHD	IAPQVTFPVD	QTVKAYGKNS	FNWVGPIPRV	NIMNPEDLKD	120
VLTKNVDFVK	PISNPLIKLL	ATGIAIYEGE	KWKHRRRIIN	PTFHSERLKR	MLPSFHQSCN	180
EMVKEWESLV	SKEGSSCELD	VWPFLENMSA	DVISRTAFGT	SYKKGQKIFE	LLREQVIYVT	240
KGFQSFYIPG	WRFLPTKMNK	RMNEINEEIK	GLIRGIIIDR	EQIIKAGEET	NDDLGLALME	300
SNLKDIREHG	KNNKNVGMMSI	EDVIQECKLF	YFAGQETTSV	LLAWTMVLLG	QNQNWDQDRAR	360
QEVQLQVFGSS	KPDFDGLAHL	KVVTMILLEV	LRLYPPVIEL	IRTIHKKTQL	GKLSLPEGVE	420
VRLPFLLIHH	DKELWGGDAN	QFNPERFSEG	VSKATKNRSL	FFPFGAGPRI	CIGQNFMSME	480
AKLALALILQ	HFTFELSPSH	AHAPSHRITL	QPQYGVRIIL	HRR		523

## SEQ ID NO:83

MENKTETTVR	RRRRIILFPV	PFQGHINPIL	QLANVLYSKG	FSITIFHTNF	NKPKTSNYPH	60
FTFRFILDND	QDERISNLP	THGPLAGMRI	PIINEHGADE	LRRELELLML	ASEEDEEVSC	120
LITDALWYFA	QSVADSLNLR	RLVMTSSLF	NFAHVSPLPQ	FDELGYLDPD	DKTRLEEQAS	180
GFPMLKVVDI	KSAYSNWQIL	KEILGKMIKQ	TKASSGVIWN	SFKELEESL	ETVIREIPAP	240
SFLIPLPKHL	TASSSSLLDH	DRTVFQWLDQ	QPPSSVLYVS	FGSTSEVDEK	DFLEIARGLV	300
DSKQSFLWV	RPGFVKGSTW	VEPLPDGFLG	ERGRIVKWVP	QQEVLAHGAI	GAFTWTHSGWN	360
STLESVCEGV	PMIFSDFGLD	QPLNARYMSD	VLKVGVIYEN	GWERGEIANA	IRRMVMDDEG	420
EYIRQNARVL	KQKADVSLMK	GGSSYESLES	LVSYISSL			458

## SEQ ID NO:84

MDAMATTEKK	PHVIFIPFPA	QSHIKAMLKL	AQLLHHKGLQ	ITFVNTDFIH	NQFLESSGPH	60
CLDGAPGFRE	ETIPDGVSHS	PEASIPPIRES	LLRSIETNFL	DRFIDLVTKL	PDPPTCIISD	120
GFLSVFTIDA	AKKLGIPVMM	YWTLAACGFM	GFYHIHSLIE	KGFAPLKDAS	YLTNGYLDTV	180
IDWVPGMEGI	RLKDFPLDWS	TDLNDKVLMF	TTEAPQSRSHK	VSHHIFHTFD	ELEPSIIKTL	240
SLRYNHIYTI	GPLQLLLDQI	PEKKQTGIT	SLHGYSLVKE	EPECFQWLQS	KEPNSVVYVN	300
FGSTTVMSLE	DMTEFGWGLA	NSNHYFLWII	RSNLVIGENA	VLPPLEEEHI	KKRGFIASWC	360
SQEKVLKHPS	VGGFLTHCGW	GSTIESLSAG	VPMICWPYSW	DQLTNCRYIC	KEWEVGLEMG	420
TKVKRDEVKR	LVQELMGE	HKMRNKAKDW	KEKARIAIAP	NGSSSLNIDK	MVKEITVLAR	480

## SEQ ID NO:85

MATSDSIVDD	RKQLHVATFP	WLAFGHILPY	LQLSKLIAEK	GHKVSFLSTT	RNIQRLSSHI	60
SPLINVVQLT	LPRVQELPED	AEATTDVHPE	DIPYLKASD	GLQPEVTRFL	EQHSPDWIY	120
DYTHYWLPSI	AASLGISRAH	FSVTTPWAIA	YMGPSADAMI	NGSDGRITVE	DLTTPKWF	180

FPTKVCWRKH	DLARLVPYKA	PGISDGYRMG	LVLKGSDDL	SKCYHEFGTQ	WLPLLETLHQ	240
VPVVPVGLLP	PEIPGDEKDE	TWVSIKKWLD	GKQKGSVVYV	ALGSEVLVSQ	TEVVELALGL	300
ELSGLPFVWA	YRKPKGPAKS	DSVELPDGFV	ERTDRGLVW	TSWAPQLRIL	SHEVCGFLT	360
HCGSGSIVEG	LMFGHPLIML	PIFGEIPRNE	EDGCLTKESV	ARSLRSVVVE	KEGEIYKANA	420
RELSKIYNDT	KVEKEYVSQF	VDYLEKNARA	VAIDHES			457

## SEQ ID NO:86

MDSGYSSSYA	AAAGMHVVIC	PWLAFGHLLP	CLDLAQRLAS	RGHRVSFVST	PRNISRLPPV	60
RPALAPLVAF	VALPLPRVEG	LPDGAESTND	VPHDRPDMVE	LHRRAFDGLA	APFSEFLGTA	120
CADWVIVDVF	HHWAAAAALE	HKVPCAMMLL	GSAHMIASIA	DRRLERAETE	SPAAAGQGRP	180
AAAPTFEVAR	MKLIRTKGSS	GMSLAERFSL	TLSSSSLVVG	RSCVEFEPET	VPLLSTLRGK	240
PITFLGLMPP	LHEGRREDGE	DATVRWLDAQ	PAKSVVYVAL	GSEVPLGVEK	VHELALGLEL	300
AGTRFLWALR	KPTGVSDADL	LPAGFEERTR	GRGVVATRWV	PQMSILAHAA	VGAFLTHCGW	360
NSTIEGLMFG	HPLIMLPIFG	DQGPNA RLIE	AKNAGLQVAR	NDGDGSFDRE	GVAAAIRAVA	420
VEEESKVFQ	AKAKKLQEI	ADMACHERYI	DGFIQQLRSY	KD		462

## SEQ ID NO:87

MSSSSSSSTS	MIDLMAAIK	GEPVIVSDPA	NASAYESVAA	ELSSMLIENR	QFAMIVTTSI	60
AVLIGCIVML	VWRRSGSGNS	KRVEPLKPLV	IKPREEEIDD	GRKKVTIFFG	TQTGTAEFGA	120
KALGEEAKAR	YEKTRFKIVD	LDDYAADDD	YEEKLKKEDV	AFFFLATYGD	GEPTDNAARF	180
YKWFTEGNDR	GEWLKLNLYG	VFGLGNRQYE	HFNKVAKVVD	DILVEQGAQR	LVQVGLGDDD	240
QCIEDDDTAW	REALWPELDT	ILREEGDTAV	ATPYTAAVLE	YRVSIHDS	AKFNDITLAN	300
GNGYTVFDAQ	HPYKANVAVK	RELHTPESDR	SCIHLEFDIA	GSGLTMKLG	HVGVLCDNLS	360
ETVDEALRL	DMSPDYFSL	HAEKEDGTPI	SSSLPPFPFP	CNLRALTALTRY	ACLLSSPKKS	420
ALVALAAHAS	DPTAERLKH	LASPAKDEY	SKWVVSQRS	LLEVMAEFPS	AKPPLGVFFA	480
GVAPRLQPRF	YSISSSPKIA	ETRIHVTCAL	VYEKMPTGRI	HKGVCSTWMK	NAVPEKSEK	540
LFLGRPIFVR	QSNFKLPDS	KVPIIMIGPG	TGLAPFRGFL	QERLALVESG	VELGPSVLFF	600
GCRNRRMDFI	YEEELQRFVE	SGALAELSVA	FSREGPTKEY	VQHKMMDKAS	DIWNMISQGA	660
YLYVCGDAKG	MARDVHRLH	TIAQEQQSMD	STKAEGFVK	LQTSGRYL	VW	712

## SEQ ID NO:88

MATSDSIVDD	RKQLHVATFP	WLAFGHILPY	LQLSKLIAEK	GHKVSFLSTT	RNIQRLSSHI	60
SPLINVQLT	LPRVQELPED	AEATTDVHPE	DIPYLKASD	GLQPEVTRFL	EQHSPDWIY	120
DYTHYWLPSI	AASLGISRAH	FSVTPWAIA	YMGPSADAMI	NGSDGRITVE	DLTTPPKWFP	180
FPTKVCWRKH	DLARLVPYKA	PGISDGYRMG	MVLKGSDDL	SKCYHEFGTQ	WLPLLETLHQ	240
VPVVPVGLLP	PEIPGDEKDE	TWVSIKKWLD	GKQKGSVVYV	ALGSEALVSQ	TEVVELALGL	300
ELSGLPFVWA	YRKPKGPAKS	DSVELPDGFV	ERTDRGLVW	TSWAPQLRIL	SHEVCGFLT	360
HCGSGSIVEG	LMFGHPLIML	PIFGDQPLNA	RLLEDKQVGI	EIPRNEEDGC	LTKEVSARS	420
RSVVVEKEGE	IYKANARELS	KIYNDTKVEK	EYVSQFVDYL	EKNARAVAI	HES	473

## SEQ ID NO:89

atggctactt	ctgattccat	cgttgacgat	agaaagcaat	tgcattgttc	tacttttcca	60
tgggttgctt	tcggctcatat	tttgccat	ttgcaattgt	ccaagttgat	tgctgaaaag	120
gggtcacagg	tttctattct	gtctaccacc	agaaacatcc	aaagattgtc	ctctcatatc	180
tccccattga	tcaacgttgt	tcaattgact	ttgccaaagag	tccaagaatt	gccagaagat	240
gctgaagcta	ctactgatgt	tcatccagaa	gatattccctt	acttgaaaaa	ggcttccgat	300
gggtttacaac	cagaagttac	tagattcttg	gaacaacatt	ccccagattg	gatcatctac	360
gattatactc	attactgggt	gccatccatt	gctgcttcatt	tggttatttc	tagagcccat	420
ttctctgtta	ctactccatg	ggctattgct	tatatgggtc	catctgctga	tgctatgatt	480
aacggttctg	atggtagaac	taccgttgaa	gatttgacta	ctccaccaa	gtggtttcca	540
tttccacaaa	aagtctgttg	gagaaaaacac	gatttggtc	gatttggtcc	atacaagct	600
ccaggtattt	ctgatgggta	cagaatgggt	atgggtttga	aaggttccga	ttgcttgttg	660
tctaagtgtc	atcatgaatt	cggtactcaa	tggttgctt	tggttgaaac	attgcatcaa	720
gttccagttg	ttccagtagg	ttgttgcca	ccagaaattc	caggtgacga	aaaagacgaa	780
acttggtgtt	ccatcaaaaa	gtggttggt	ggtaagcaaa	agggttctgt	tgtttatgtt	840
gctttgggtt	ccgaagcttt	ggtttctcaa	accgaagtgt	ttgaattggc	tttgggtttg	900
gaattgtctg	gtttgcat	tggttggtg	taacagaaaac	ctaaaggtcc	agctaagtct	960
gattctgttg	aattgccaga	tggtttcggt	gaaagaacta	gagatagagg	tttgggtttg	1020
acttctgttg	ctccacaatt	gagaattttg	tctcatgaat	ccgtctgttg	tttcttgact	1080
cattgtggtt	ctggttctat	cgttggaagt	ttgatgtttg	gtcaccatt	gattatgttg	1140

ccaatctttg	gtgaccaacc	attgaacgct	agattattgg	aagataagca	agtcgggtatc	1200
gaaatcccaa	gaaatgaaga	agatgggtgc	ttgaccaaag	aatctgttgc	tagatctttg	1260
agatccgttg	tcgttgaaaa	agaagggtgaa	atctacaagg	ctaacgctag	agaattgtcc	1320
aagatctaca	acgataccaa	ggtcgaaaaa	gaatacgttt	cccaattcgt	tgactacttg	1380
gaaaagaatg	ctagagctgt	tgccattgat	catgaatctt	ga		1422

## SEQ ID NO:90

atggaagctt	ctagagcatc	ttgtgttget	ttgtgtgttg	tttgggtttc	catcgttatt	60
actttggctt	ggagagtttt	gaattgggtc	tggttaagac	caaaaaagtt	ggaaagatgc	120
ttgagagaac	aagggtttgac	tggttaactct	tacagattgt	tggttcggtga	taccaaggac	180
ttgtctaaag	tggttgaaac	aactcaatcc	aagcctatca	agttgtctac	ctctcatgat	240
attgtctcaa	gagttactcc	attcttccat	agaactgtta	actccaacgg	taagaactct	300
tttgtttgga	tggttccaat	tccaagagtc	catattatga	accctgaaga	tttgaaggac	360
gctttcaaca	gacatgatga	tttccataag	accgtcaaga	acccaattat	gaagtctcca	420
ccaccaggtg	tagttgggtat	tgaagggtgaa	caatgggcca	aacatagaaa	gattattaac	480
ccagccttcc	acttggaata	gttgaaagggt	atggttccaa	tcttctacca	atcctgctct	540
gaaatgatta	acaagtggga	atccttgggt	tccaaagaat	cttctctgtga	attggatgtc	600
tggtccatatt	tggaataact	cacctccgat	gttattttcca	gagctgcttt	tggttcttct	660
tacgaagaag	gtagaaagat	cttccaatta	ttgagagaag	aagccaagggt	ttactccgtt	720
gctttgagat	ctgtttacat	tccaggttgg	agattcttgc	caactaagca	aaacaaaaag	780
accaaaagaaa	tccacaacga	aatcaagggt	ttgttgagg	gtatcatcaa	caagagagaa	840
gaagctatga	aggctggtga	agctacaaaa	gatgatttgt	tggttatctt	gatggaatcc	900
aacttcagag	aaatccaaga	acacggtaac	aacaagaatg	ccggtatgtc	tattgaagat	960
gttatcgggt	aatgcaagtt	gttctacttt	gctgggtcaag	aaactacctc	cgttttgttg	1020
gtttggacca	tgattttgtt	gtcccaaat	caagattggc	aagctagagc	tagagaagaa	1080
gtcttgaaag	ttttcgggtc	taacatccca	acctacgaag	aattgtctca	cttgaagggt	1140
gtcactatga	tcttgttgga	agtattgaga	ttatacccat	ccgttgttgc	attgccaaga	1200
actactcata	agaaaaactca	attgggtaaa	ttgtccttgc	cagctggtgt	tgaagtttct	1260
ttgccaattt	tggttagtcca	ccacgacaaa	gaattgtggg	gtgaagatgc	taatgaattc	1320
aagccagaaa	gatttctcga	agggtgttct	aaagctacca	agaacaagtt	cacttacttg	1380
ccatttgggt	gtgggtccaag	aatatgtatt	ggtcaaaaatt	tcgctatggt	cgaagctaaa	1440
ttggcttttg	ctttgatctt	gcaacatttc	gctttcgaat	tgtcaccatc	ttatgctcat	1500
gctccatctg	ctgttattac	attgcaacca	caatttgggt	cccatatcat	cttgcataag	1560
agataac						1567

## SEQ ID NO:91

MEASRASCVA	LCVVVWSIVI	TLAWRVLNWV	WLRPKKLERC	LREQGLTGNS	YRLIFGDTKD	60
LSKMLEQTQS	KPIKLSSTSHD	IAPRVTPFFH	RTVNSNGKNS	FVWMGPPIRV	HIMNPEDLKD	120
AFNRHDDFHK	TVKNPIMKSP	PPGIVGIEGE	QWAKHRKIIN	PAFHLEKLKG	MVPIFYQSCS	180
EMINKWESLV	SKESSCELDV	WPLYENFTSD	VISRAAFGSS	YEGRKIFQL	LREEAKVYSV	240
ALRSVYIPGW	RFLPTKQNK	TKEIHNEIKG	LLKGIINKRE	EAMKAGEATK	DDLGLILMES	300
NFREIQEHGN	NKNAGMSIED	VIGECKLFYF	AGQETTSVLL	VWTMILLSQN	QDWQARAREE	360
VLKVFSGSNIP	TYEELSHLKV	VTMILLEVL	LYPSVVALPR	TTHKKTQLGK	LSLPAGVEVS	420
LPILLVHDK	ELWGEDANEF	KPERFSEGV	KATKNKFTYL	PFGGGRIC	QNFAMVEAK	480
LALALILQHF	AFELSPSYAH	APSAVITLQP	QFGAHILHK	R		521

## SEQ ID NO:92

ASWVAVLSVV	WVSMVIAWAW	RVLNVWVLRP	KKLEKCLREQ	GLAGNSYRLL	FGDTKDLSKM	60
LEQTQSKPIK	LSTSHDIAPH	VTPFFHQTVN	SYGKNSFVWM	GPIPRVHIMN	PEDLKDTFNR	120
HDDFHKVVKV	PIMKSLPQGI	VGIEGEQWAK	HRKIINPAFH	LEKLKGMVPI	FYRSCSEMIN	180
KWESLVSKES	SCELDVWPYL	ENFTSDVISR	AAFSSYEEG	RKIFQLLREE	AKIYTVAMRS	240
VYIPGWRFLP	TKQNKKAKEI	HNEIKGLLKG	IINKREEAMK	AGEATKDDL	GILMESNFRE	300
IQEHGNNKNA	GMSIEDVIGE	CKLFYFAGQE	TTSVLLVWTM	VLLSQNDWQ	ARAREEVLOV	360
FGSNIPTYEE	LSQLKVVTMI	LLEVLRLYPS	VVALPRTHK	KTQLGKLSLP	AGVEVSLPIL	420
LVHDKELWG	EDANEFKPER	PSEGVSKATK	NQFTYFFFG	GPRICIGQNF	AMMEAKLALS	480
LILRHFALEL	SPLYAHAPSV	TITLQPQYGA	HIILHKR			517

## SEQ ID NO:93

MEASRPSCVA	LSVVLSIVI	AWAWRVLNWV	WLRPNKLERC	LREQGLTGNS	YRLIFGDTKE	60
ISMMEQAQS	KPIKLSSTHD	IAPRVIPFHS	QIVYTYGRNS	FVWMGPTPRV	TIMNPEDLKD	120

AFNKSDEFQR	AINSPIVKSI	SQGLSSLEGE	KWAKHRKIIN	PAFHLEKLKG	MLPTFYQSCS	180
EMINKWESLV	FKEGSREMDV	WPLYENLTSD	VISRAAFGSS	YEEGRKIFQL	LREEAKFYTI	240
AARSVYIPGW	RFLPTKQNK	MKEIHKEVRG	LLKGIINKRE	DAIKAGEAAK	GNNLGILMES	300
NFREIQEHGN	NKNAGMSIED	VIGECKLFYF	AGQETTSVLL	VWTLVLLSON	QDWQARAREE	360
VLQVFGTNIP	TYDQLSHLKV	VTMILLEVL	LYPAVVELPR	TTYKKTQLGK	FLLPAGVEVS	420
LHIMLAHHDK	ELWGEDAKEF	KPERFSEGV	KATKNQFTYF	PFGAGPRICI	GQNFAMLEAK	480
LALSILQHF	TFELSPSYAH	APSVTITLHP	QFGAHFILHK	R		521

## SEQ ID NO:94

CVALSIVLV	IVIAWAWRL	NWVWLRPNKL	ERCLREQGLT	GNSYRLLFGD	TKEISMMVEQ	60
AQSKPIKLST	THDIAPRVIP	FSHQIVYTYG	RNSFVWMGPT	PRVTIMNPED	LKDAFNKSD	120
FQRAISNPIV	KSISQGLSSL	EGEKWAKHRK	IINPAFHLEK	LKGMLPTFYQ	SCSEMINKWE	180
SLVFEKGSRE	MDVWPLYENL	TSDIVISRAAF	GSSYEGRKI	FQLLREEAKF	YTIAARSVYI	240
PGWRFLPTKQ	NKRMEIHK	VRGLLKGIIIN	KREDIAKAGE	AAKGNLLGIL	MESNFREIQE	300
HGNNKNAGMS	IEDVIGECKL	FYFAGQETTS	VLLVWTLVLL	SONQDWQARA	REEVLQVFGT	360
NIPTYDQLSH	LKVVTMILLE	VLRLYPVAVVE	LPRTTYKKTQ	LKGKLLPAGV	EVSLHIMLAH	420
HDKELWGEDA	KEFKPERFSE	GVSKATKNQF	TYFFPGAGPR	ICIGQNFAML	EAKLALSIL	480
QHFTFELSPS	YAHAPSVTIT	LHPQFGAHFI	LHKKR			514

## SEQ ID NO:95

MGPIPRVHIM	NPEDLKDTFN	RHDDFHKVVK	NPIMKSLPQG	IVGIEGDQWA	KHRKIINPAF	60
HLEKLKGMVP	IFYQSCSEMI	NIWKSLSVSKE	SSCELDVWPY	LENFTSDVIS	RAAFGSSYEE	120
GRKIFQLLRE	EAKVYTVAVR	SVYIPGWREL	PTKQNKKTKE	IHNEIKGLLK	GIINKREEAM	180
KAGEATKDDL	LGILMESNFR	EIQEHGNNKN	AGMSIEDVIG	ECKLFYFAGQ	ETTSVLLVWT	240
MVLLSQNQDW	QARAREEVLQ	VFGSNIPTYE	ELSHLKVVTM	ILLEVLRLLYP	SVVALPRTHH	300
KKTQLGKLSL	PAGVEVSLPI	LLVHHDKELW	GEDANEFKPE	RFSEGVSKAT	KNQFTYFPFG	360
GGPRICIGQN	FAMMEAKLAL	SLILQHFTFE	LSPQYSHAPS	VTITLQPQYG	AHLILHKKR	418

## SEQ ID NO:96

atggaagcat	caagggtctag	ttgtgttgctg	ctatgtgttg	tttgggtgag	catagtaatt	60
acattggcat	ggagggtgct	gaattgggtg	tggttgaggc	caaagaaact	agaaagatgc	120
ttgagggagc	aaggccttac	aggcaattct	tacaggcttt	tggttgagga	caccaaggat	180
ctctcgaaaga	tgctggaaca	aacacaatcc	aaacccatca	aactctccac	ctcccatgat	240
atagcgccac	gagtcacccc	atttttccat	cgaactgtga	actctaattg	caagaattct	300
ttgttttga	tgggccctat	accaagagt	cacatcatga	atccagaaga	tttgaaagat	360
gccttcaaca	gacatgatga	tttccataag	acagtataaa	atcctatcat	gaagtctcca	420
ccaccgggca	ttgtaggcat	tgaagggtgag	caatgggcta	aacacagaaa	gattatcaac	480
ccagcattcc	atttagagaa	gctaaagggt	atggtaccaa	tattttacca	aagttgtagc	540
gagatgatta	acaaatggga	gagcttggtg	tccaaagaga	gttcattgtga	gttggtatgtg	600
tggccttatc	ttgaaaaatt	taccagcgat	gtgatttccc	gagctgcatt	tggaaagtagc	660
tatgaagagg	gaaggaaaat	atttcaacta	ctaagagagg	aagcaaaagt	ttattcggtta	720
gctctacgaa	gtgtttacat	tccaggatgg	aggtttctac	caaccaagca	gaacaagaag	780
acgaaggaaa	ttcaccaatga	aattaaaggc	ttacttaagg	gcattataaa	taaaagggaa	840
gaggcgatga	aggcagggga	agccactaaa	gatgacttac	taggaatact	tatggagtcc	900
aacttcaggg	aaattcagga	acatgggaac	aacaaaaatg	ctggaatgag	tattgaagat	960
gtaattggag	agtgttaagt	gttttacttt	gctgggcaag	agaccacttc	ggtgttgctt	1020
gttttgacaa	tgatttttact	aagccaaaaat	caggattggc	aagctcgtgc	aagagaagag	1080
gtcttgaaag	tccttggaag	caacatccca	acctatgaag	agctaagtca	cctaaaagtt	1140
gtgaccatga	ttttacttga	agttcttcga	ttatacccat	cagtcgttgc	gcttccctga	1200
accactcaca	agaaaaacaca	gcttggaaaa	ttatcattac	cagctggagt	ggaagtctcc	1260
ttgcccatac	tgcttggttca	ccatgacaaa	gagttgtggg	gtgaggatgc	aaatgagttc	1320
aagccagaga	ggtttttcaga	gggagtttca	aaggcaacaa	agaacaaatt	tacatactta	1380
cctttcggag	ggggtccaag	gatttgcatt	ggacaaaact	ttgccatggt	ggaagctaaa	1440
ttggccttgg	ccctgatttt	acaacacttt	gcctttgagc	tttctccatc	ctatgctcat	1500
gctccttctg	cagttataac	ccttcaacct	caatttggtg	ctcatatcat	tttgcataaa	1560
cgttga						1566

## SEQ ID NO:97

atgtcctcca	actccgattt	ggtcagaaga	ttggaatctg	ttttgggtgt	ttctttcgtg	60
------------	------------	------------	------------	------------	------------	----

ggttctgtta	ctgattccgt	tggtgttatt	gctaccacct	ctattgcttt	ggttatcggt	120
gttttggtt	tggtgtggag	aagatcctct	gacagatcta	gagaagttaa	gcaattggct	180
gttccaaagc	cagttactat	cggtgaagaa	gaagatgaat	tcgaagtgc	ttctggtaag	240
accagagttt	ctattttcta	cggtactcaa	actggtagctg	ctgaagggtt	tgctaaggct	300
ttggctgaag	aaatcaaagc	cagatacgaa	aaagctgccg	ttaagggtat	tgatttggat	360
gattacacag	ccgaagatga	caaatcacgt	gaaaagtga	agaaagaaac	tatggccttc	420
ttcatgttgg	ctacttatgg	tgatggtgaa	cctactgata	atgctgctag	attttacaag	480
tggttcaccg	aaggtagctga	tagaggtggt	tggttggaac	atttgagata	cggtgtattc	540
ggtttggtga	acagacaata	cgaacacttc	aacaagattg	ccaaggttgt	tgatgatttg	600
ttggttgaac	aaggtgccaa	gagattgggt	actggtgggt	tggttgatga	tgatcaatgc	660
atcgaagatg	atttctccgc	ttggaaagaa	gccttggtgc	cagaattgga	tcaattattg	720
caagatgata	ccaacacogt	ttctactcca	tacactgctg	ttattccaga	atacagagtt	780
gttatccacg	atccatctgt	tacctcttat	gaagatccat	actctaact	ggctaaccgt	840
aatgcctctt	acgatattca	tcattccatgt	agagctaaccg	ttgccgtcca	aaaagaattg	900
cataagccag	aatctgacag	aagttgcatc	catttggaat	tcgatatttt	cgctactggt	960
ttgacttacg	aaaccgggtga	tcattgttgt	gtttacgctg	ataattgtga	tgatactgta	1020
gaagaagccg	ctaagtgtgt	gggtcaacca	ttggatttgt	tgttctccat	tcataccgat	1080
aacaacgacg	gtacttcttt	gggttcttct	ttgccaccac	catttccagg	tccatgtact	1140
ttgagaactg	ctttggctag	atatgccgat	ttgttgaatc	caccaaaaaa	ggctgctttg	1200
attgctttag	ctgctcatgc	tgatgaacca	tctgaagctg	aaagattgaa	gttcttgtca	1260
tctccacaag	gtaaggacga	atattctaaa	tggtgtgtcg	gttcccaaag	atccttggtt	1320
gaagttagtg	ctgaatttcc	atctgctaaa	ccaccattgg	gtgtattttt	tgctgctggt	1380
gttccctagat	tgcaaccttc	atattactcc	atctcttcca	gtccaagatt	tgctccacat	1440
agagttcatg	ttacttgctg	gttcccaactc	caactggtag	aattcacaga		1500
ggtgtatggt	cattctggat	gaagaatggt	gtcccattgg	aaaagtctca	aaactgttct	1560
tggtccccaa	ttttcatcag	acaatctaatt	ttcaagttgc	cagccgatca	ttctgttcca	1620
atagttatgg	ttggtccagg	tactggttta	gtcctcttta	gaggtttctt	acaagaaga	1680
ttggccttga	aagaagaagg	tgctcaagtt	ggctcctgctt	tggtgttttt	tggttgacga	1740
aacagacaaa	tggaacttcat	ctacgaagtc	gaattgaaca	actttgtcga	acaagtgct	1800
ttgtccgaat	tgatcggtgc	tttttcaaga	gaaggtccat	ccaaagaata	cgtccaacat	1860
aagatgggtg	aaaaggcagc	ttacatgtgg	aacttgattt	ctcaagggtg	ttacttctac	1920
gtttgtggtg	atgctaaagg	tatggctaga	gatgttcata	gaacattgca	taccatcgtc	1980
caacaagaag	aaaagggtga	ttctaccaag	gccgaatcca	tcgttaagaa	attgcaaatg	2040
gacggtagat	acttgagaga	tggttggtga				2070

## SEQ ID NO:98

MSSNSDLVRR	LESVLGVSG	GSVTDVVVI	ATTSIALVIG	VLVLLWRRSS	DRSREVKQLA	60
VPKPVITVEE	EDEFVVASGK	TRVSIFYGTQ	TGTAEGFAKA	LAEEIKARYE	KAQVKVIDLD	120
DYTAEDDKYG	EKLKKTMAF	FMLATYGDGE	PTDNAARFYK	WFTGTDGRV	WLEHLRYGVF	180
GLGNRQYEHF	NKIAKVDDL	LVEQGAQRLV	TVGLGDDQD	IEDDFS AWKE	ALWPELDQLL	240
QDDTNTVSTP	YTAVIPEYRV	VIHDPVTSY	EDPYSNMANG	NASYDIHHP	RANVAVQKEL	300
HKPESDRSCI	HLEFDIFATG	LYETGDHVG	VYADNCDDTV	EAAKLLGQP	LDLLFSIHTD	360
NNDGTSLGSS	LPPPFPGPCT	LRTALARYAD	LLNPPKKAAL	IALAAHADEP	SEAERLKFLS	420
SPQKGDEYSK	WVVGSRSLV	EVMAEFPSAK	PPLGVFFAAV	VPRQLPRYYS	ISSSPRFAPH	480
RVHVTALVY	GPTPTGRIHR	GVCSFWMKNV	VPLEKSQNC	WAPIFIRQSN	FKLPADHSVP	540
IVMVGPGTGL	APFRGFLQER	LALKEEGAQV	GPALLFFGCR	NRQMDFIYEV	ELNNFVEQGA	600
LSELIVAFSR	EGPSKEYVQH	KMVEKAAYMW	NLISQGGYFY	VCGDAKGMAR	DVHRTLHTIV	660
QEEKVDSSTK	AESIVKKLQM	DGRYL RDVW				689

## SEQ ID NO:99

atggatgctg	tgacgggttt	gttaactgtc	ccagcaaccg	ctataactat	tggtggaact	60
gctgtagcat	tgccggtagc	gctaactctt	tggtacctga	aatcctacac	atcagctaga	120
agatcccaat	caaatcatct	tccaagagtg	cctgaagtcc	caggtgttcc	attgttagga	180
aatctgttac	aattgaagga	gaaaaagcca	tacatgactt	ttacgagatg	ggcagcgaca	240
tatggaccta	tctatagtat	caaaaactggg	gctacaagta	tggttggtgt	atcatctaatt	300
gagatagcca	aggaggcatt	ggtgaccaga	ttccaatcca	tatctacaag	gaacttatct	360
aaagccctga	aagtacttac	agcagataag	acaatggctg	caatgtcaga	ttatgatgat	420
tatcataaaa	cagttaaagag	acacatactg	accgccgtct	tggttccctaa	tgacacagaaa	480
aagcatagaa	ttcacagaga	tatcatgatg	gataacatat	ctactcaact	tcatgaattc	540
gtgaaaaaca	accagagaaca	ggaagaggta	gaccttagaa	aaatctttca	atctgagtta	600
ttcggcttag	ctatgagaca	agccttagga	aaggatgttg	aaagtttgta	cgttgaagac	660
ctgaaaatca	ctatgaatag	agacgaaatc	tttcaagctc	ttgttggtga	tccaatgatg	720
ggagcaatcg	atggttgattg	gagagacttc	tttccatacc	taaagtgggt	cccaaacaaa	780



aagttcgaaa	atactattca	acaaatgtac	atcagaagag	aagctgttat	gaaatcttta	840
atcaaagagc	acaaaaagag	aatagcgtca	ggcgaaaagc	taaatagtta	tatcgattac	900
cttttatctg	aagctcaaac	tttaaccgat	cagcaactat	tgatgtcctt	gtgggaacca	960
atcattgaaat	cttcagatac	aacaatgggtc	acaacagaat	gggcaatgta	cgaattagct	1020
aaaaacccta	aattgcaaga	taggttgtac	agagacatta	agtcctgtctg	tggatctgaa	1080
aagataaaccg	aagagcatct	atcacagctg	ccttacatta	cagctatttt	ccacgaaaca	1140
ctgagaagac	actcaccagt	tcctatcatt	cctctaagac	atgtacatga	agataccgtt	1200
ctaggcggct	accatgttcc	tgctggcaca	gaacttgccg	ttaacatcta	cggttgcaac	1260
atggacaaaa	acgtttggga	aaatccagag	gaatggaacc	cagaaagatt	catgaaagag	1320
aatgagacaa	ttgattttca	aaagacgatg	gccttcgggtg	gtggtaagag	agttttgtgt	1380
ggttccttgc	aagccctttt	aactgcatct	attgggattg	ggagaatgg	tcaagagttc	1440
gaatggaac	tgaaggat	gactcaagag	gaagtgaaca	cgataggcct	aactacacaa	1500
atgttaagac	cattgagagc	tattatcaaa	cctaggatcc	catcaagacc	aagtcctagt	1560
accgaacaat	ctgcaaaaaa	agttagaaaa	aaagcagaaa	atgcacacaa	tactccattg	1620
ctagttcttt	atggttctaa	tatgggaaca	gcggaaggaa	cggccaggga	tctagctgac	1680
atagctatgt	ccaagggatt	tgccccgcaa	gtagcaaccc	tggattccca	tgaggtaac	1740
ttgccaagag	aaggtgctgt	tctaatagtt	accgctagct	acaatgggca	ccctccagat	1800
aatgcgaagc	agttcgtcga	ttggttagat	caagcatcag	cagatgaagt	taagggtgtt	1860
agatactctg	tttttgatg	tgagataag	aattgggcca	ccacatatca	gaaggttccg	1920
gctttcatcg	atgaaatgct	tgctgcaaaa	ggggctgaaa	atatagcaga	tcgtggtgag	1980
gccgacgcaa	gcgacgattt	tgaggggtacc	tatgaggagt	ggagagagca	catgtggtct	2040
gattgttgcg	cgtattttta	cttagacata	gaaaattctg	aagacaataa	aagtgccctta	2100
cttcttcaat	tcgtcgatag	tgctgctggac	atgcccttag	caaagatgca	tggagccttt	2160
tcaacgaacg	tagtagccag	taaggaactt	caacaaccag	gtagtgccag	aagtacacgt	2220
cacttggaag	ttgaattacc	aaaagaggca	tcctaccaag	aaggtgacca	tcttggtgta	2280
atcccaagaa	actacgaagg	tatagtcaat	agggtaacgg	caagatttgg	gctggatgca	2340
agccaacaga	taagactaga	agcagaagaa	gaaaaattgg	cgcaccttcc	actagcgaag	2400
acagtatccg	ttgaagaatt	attgcaatac	gtggaattgc	aggatcccgt	cactagaacg	2460
caattgagag	ctatggcagc	aaagactgtt	tgccacctc	acaagggtga	acttgaagct	2520
ctacttgaaa	aacaagcata	caaagagcaa	gtgctagcaa	agagactaac	catgttagaa	2580
ttgctggaaa	aatacccggc	atgcgaaatg	gaattctccg	aatttatcgc	gttggtgcca	2640
agtattcgtc	ccaggtatta	ctcaatttca	tcttcaccaa	gggttgacga	gaaacaggca	2700
tctattaccg	tatctgttgt	ctctggagaa	gcttgagtg	gttacggaga	atacaagggt	2760
attgcttcca	attatcttgc	agaactgcag	gaaggggata	caattacctg	cttattttct	2820
actcctcaat	cagaatttca	tcttccgaag	gatccagaaa	ctccgttaat	tatggtaggt	2880
ccgggaacag	gagtcgcccc	tttcagagcg	tttgtgcaag	caaggaagca	actaaaagaa	2940
cagggacaaa	gtctgggtga	ggcacatcla	tatttcggtt	gcagatctcc	gcatgaggat	3000
tacttatacc	aagaagaact	tgaaaacgcc	caatcagaag	gtattatcac	cttgcatact	3060
gcattcagta	gaatgccaaa	ccagccgaaa	acttacgtac	agcatgttat	ggagcaagat	3120
gttaagaagt	taattgagct	tttgataag	ggcgccact	tctacatttg	cggcgacgga	3180
gtccaaatgg	cgcctgccgt	tgaagccacc	ttgatgaaat	catatgcaga	tgttcatcaa	3240
gtttcagaag	cggacgcccc	tctttggtta	caacaactag	aggagaaagg	aaggtatgca	3300
aaagatgttt	ggtaa					3315

## SEQ ID NO:100

MDAVTGLLTV	PATAITIGGT	AVALAVALIF	WYLKSYTSAR	RSQSNHLPRV	PEVPGVPLL	60
NLLQLKEKKP	YMTFTRWAAT	YGPYISIKTG	ATSMVVVSSN	EIAKEALVTR	FQSISTRNLS	120
KALKVLTADK	TMVAMSDYDD	YHKTVKRHIL	TAVLGPNAQK	KHRIHRDIMM	DNISTQLHEF	180
VKNNPEQEEV	DLRKIFQSEL	FGLAMRQALG	KDVESLYVED	LKITMNRDEI	FQVLVVDPM	240
GAIDVDWRDF	PPLYKWPVNK	KFENTIQQMY	IRREAVMKSL	IKEHKKRIAS	GEKLNSYIDY	300
LLSEAQTLTD	QQLMLSLWEP	IISSDITMVI	TTEWAMYELA	KNPKLQDRLY	RDIKSVCGSE	360
KITEEHLSQL	PYITAI FHET	LRRHSPVPII	PLRHVHEDTV	LGGYHVPAGT	ELAVNIYGCN	420
MDKNVWENPE	EWNPERFMKE	NETIDFQKTM	AFGGGKRVCA	GSLQALLTAS	IGIGRMVQEF	480
EWKLKDMTQE	EVNTIGLTQ	MLRPLRAIK	PRIPSRPSPS	TEQSAKKVRK	KAENAHNTPL	540
LVLYGSNMGT	AEGTARDIAD	IAMSKGFAPQ	VATLD SHAGN	LPREGAVLIV	TASYNHPPD	600
NAKQFVDWLD	QASADEVKGV	RYSVEGCGDK	NWATTYQKVP	AFIDEMLAAK	GAENIADRGE	660
ADASDDFEGT	YEEWREHMWS	DVAAYFNLDI	ENSEDNKSAL	LLQFVDSAAD	MPLAKMHGAF	720
STNVVASKEL	QPPGSARSTR	HLIELPKEA	SYQEGDHLGV	IPRNYEGIVN	RVTARFGLDA	780
SQQIRLEAEE	EKLAHLPLAK	TVSVEELLQY	VELQDPVTRT	QLRAMAAKT	CPPHKVELEA	840
LLEKQAYKEQ	VLAKRLLTME	LLEKYPACEM	EFSEFIALLP	SIRPRYSIS	SSPRVDEKQA	900
SITVSVVSGE	AWSGYGEYKG	IASNYLAELQ	EGDTITCFIS	TPQSEFTLPK	DPETPLIMVG	960
PGTGVAPFRG	FVQARKQLKE	QGQSLGEAHL	YFGCRSPHED	YLYQEELENA	QSEGIITLHT	1020
AFSRMPNQPK	TYVQHVMEQD	GKKLIELLDK	GAHFYICGDG	SQMAPAVEAT	LMKSYADVHQ	1080
VSEADARLWL	QQLEEKGRYA	KDWW				1104

## SEQ ID NO:101

atggatgctg	tgacggggtt	gttaactgtc	ccagcaaccg	ctataactat	tgggtggaact	60
gctgtagcat	tggcggtagc	gctaactctt	tggtagctga	aatcctacac	atcagctaga	120
agatcccaat	caaatacatc	tccaagagtg	cctgaagtcc	cagggtgtcc	attgttagga	180
aatctgttac	aattgaagga	gaaaaagcca	tacatgaact	ttacgagatg	ggcagcgaca	240
tatggaccta	tctatagtat	caaaaactgg	gctacaagta	tgggtgtggt	atcatcctaat	300
gagatagcca	aggaggcatt	ggtgaccaga	ttccaatcca	tatctacaag	gaacttatct	360
aaagccctga	aagtacttac	agcagataag	acaatggtcg	caatgtcaga	ttatgatgat	420
tatcataaaa	cagtttaagag	acacatactg	accgcgctct	tgggtcctaa	tgacacagaaa	480
aagcatagaa	ttcacagaga	tatcatgatg	gataacatat	ctactcaact	tcatgaattc	540
gtgaaaaaca	acccagaaca	ggaagaggta	gaccttagaa	aaatctttca	atctgagtta	600
ttcggcttag	ctatgagaca	agccttagga	aaggatgttg	aaagtttgta	cgttgaagac	660
ctgaaaaatca	ctatgaaatg	agacgaaatc	tttcaagtcc	ttgttgttga	tccaatgatg	720
ggagcaatcg	atgttgattg	gagagacttc	tttccatacc	taaagtgggt	cccaaacaaa	780
aagttcgaata	atactattca	acaaatgtac	atcagaagag	aagctgttat	gaaatcttta	840
atcaaagagc	acaaaaagag	aatagcgtca	ggcgaaaagc	taaatagtta	tatcgattac	900
cttttatctg	aagctcaaac	tttaaccgat	cagcaactat	tgatgtcctt	gtgggaacca	960
atcattgaat	cttcagatac	aacaatgggt	acaacagaat	gggcaatgta	cgaattagct	1020
aaaaacccta	aattgcaaga	taggttgtac	agagacatta	agtccgtctg	tggatctgaa	1080
aagataaaccg	aagagcatct	atcacagctg	ccttacatta	cagctatttt	ccacgaaaca	1140
ctgagaagac	actcaccagt	tcctatcatt	cctctaagac	atgtacatga	agataccgtt	1200
ctaggcggct	accatgttcc	tgctggcaca	gaacttgccg	ttaacatcta	cggttgcaac	1260
atggacaaaa	acgtttggga	aaatccagag	gaatggaacc	cagaaagatt	catgaaagag	1320
aatgagacaa	ttgattttca	aaagacgatg	gccttcgggt	gtggtaagag	agtttgtgct	1380
ggttccttgc	aagccctttt	aactgcatct	attgggattg	ggagaatggg	tcaagagttc	1440
gaatggaaac	tgaaggatat	gactcaagag	gaagtgaaca	cgataggcct	aactacacaa	1500
atgttaagac	cattgagagc	tattatcaaa	cctaggatcc	catcaagacc	aagtcctagt	1560
accgaacaat	ctgcaaaaaa	agttagaaaa	aaagcagaaa	atgcacacaa	tactccattg	1620
ctagtctctt	atggttctaa	tatgggaaca	gcggaaaggaa	cggccaggga	tctagctgac	1680
atagctatgt	ccaagggatt	tgcccgcgaa	gtagcaaccc	tggattccca	tgacaggtaac	1740
ttgccaagag	aaggtgctgt	tctaataagt	accgctagct	acaatgggca	ccctccagat	1800
aatgcgaagc	agttcgtcga	ttgggttagat	caagcatcag	cagatgaagt	taaggggtgt	1860
agataactctg	tttttggatg	tgagagataag	aattgggcca	ccacatatca	gaaggttccg	1920
gctttcatcg	atgaaatgct	tgctgcacaa	ggggctgaaa	atatagcaga	tcgtggtgag	1980
gccgacgcaa	gcgacgattt	tgagggtagc	tatgaggagt	ggagagagca	catgtgtgct	2040
gatgttgccg	cgtattttta	tctagacata	gaaaattctg	aagacaataa	aagtgcctta	2100
cttcttcaat	tcgtcgtatg	tgctgcggac	atgcccttag	caaagatgca	tggagccttt	2160
tcaacgaacg	tagtagccag	taaggaaact	caacaaccag	gtagtgccag	aagtacacgt	2220
cacttggaag	ttgaattacc	aaaagaggca	tcctaccaag	aaggtgacca	tcttggtgta	2280
atcccaagaa	actacgaagg	tatagtcaat	agggtaacgg	caagatttgg	gctggatgca	2340
agccaacaga	taagactaga	agcagaagaa	gaaaaattgg	cgcaccttcc	actagcgaag	2400
acagtatccg	ttgaagaatt	attgcaatac	gtggaattgc	aggatcccg	cactagaacg	2460
caattgagag	ctatggcagc	aaagactgtt	tgtccacctc	acaaggttga	acttgaagct	2520
ctacttgaaa	aacaagcata	caaagagcaa	gtgctagcaa	agagactaac	catgttagaa	2580
ttgctggaag	aatacccgcc	atgcgaaatg	gaattctccg	aatttatcgc	gttgttgcca	2640
agtatctcgt	ccaggtatta	ctcaatttca	tcttcacca	gggttgacga	gaaacaggca	2700
tctattaccg	tatctgtggt	ctctggagaa	gcttgaggat	gttacggaga	atacaagggt	2760
attgcttcca	attatcttgc	agaactgcag	gaaggggata	caattacctg	ctttatttct	2820
actcctcaat	cagaattttac	tcttcggaag	gatccagaaa	ctccgttaat	tatggtaggt	2880
cggggaacag	gagtcgcccc	tttcagaggc	tttgtgcaag	caaggaagca	actaaaagaa	2940
caggagacaaa	gtctgggtga	ggcacatcta	tatttcgggt	gcagatctcc	gcatagagat	3000
tacttatacc	aagaagaact	tgaaaacgcc	caatcagaag	gtattatcac	cttgcatact	3060
gcattcagta	gaatgccaaa	ccagccgaaa	acttacgtac	agcatgttat	ggagcaagat	3120
ggtaagaagt	taattgagct	tttgataag	ggcgcccaact	tctacatttg	ggcgacagga	3180
tcccgaatgg	cgctgcctg	tgaagccacc	ttgatgaaat	catatgcaga	tggtcatcaa	3240
gtttcagaag	cggagcggcg	tctttgggta	caacaactag	aggagaaagg	aaggtatgca	3300
aaagatgttg	cttaa					3315

## SEQ ID NO:102

MDAVTGLLT	PATAITIGGT	AVALAVALIF	WYLKSYTSAR	RSQSNHLPRV	FEVPGVPLLG	60
NLLQLKEKP	YMTFTRWAAT	YGPYISIKTG	ATSMVVVSSN	EIAKEALVTR	FQSISTRNLS	120
KALKVLTAADK	TMVAMSDYDD	YHKTVKRHIL	TAVLGPNAQK	KHRIHRDIMM	DNISTQLHEF	180
VKNNPQEVEV	DLRKIFQSEL	FGLAMRQALG	KDVESLYVED	LKITMNRDEI	FQVLVVDPM	240

GAIDVDWRDF	FPYLKWVPNK	KFENTIQQMY	IRREAVMKS	IKEHKKRIAS	GEKLSYIDY	300
LLSEAOQLTD	QQLMLSLWEP	IISSDSTMV	TTEWAMYELA	KNPKLQDRLY	RDIKSVCGSE	360
KITEEHLSQL	PYITAIFHET	LRRHSPVPII	PLRHVHEDTV	LGGYHVPAGT	ELAVNIYGCN	420
MDKNVWENPE	EWNPERFMKE	NETIDFQKTM	AFGGGKRVCA	GSLQALLTAS	IGIGRMVQEF	480
EWKLKDMTQE	EVNTIGLTQ	MLRPLRAIIK	PRIPSRSPSPS	TEQSAKKVRK	KAENAHNTPL	540
LVLYGSNMGT	AEGTARDLAD	IAMSKGFAPQ	VATLDSHAGN	LPREGAVLIV	TASYNGHPPD	600
NAKQFVDWLD	QASADEVKGV	RYSVFGCGDK	NWATTYQKVP	AFIDEMLAAC	GAENIADRGE	660
ADASDDFEGT	YEEWREHMWS	DVAAYFNLDI	ENSEDNKSAL	LLQFVDSAAD	MPLAKMHGAF	720
STNVVASKEL	QQPGSARSTR	HLEIELPKEA	SYQEGDHLGV	IPRNYEGIVN	RVTARFGLDA	780
SQOIRLEAEE	EKLAHLPLAK	TVSVEELLQY	VELQDPVTRT	QLRAMAAKT	CPPHKVELEA	840
LLEKQAYKEQ	VLAKRRLTME	LLEKYPACEM	EFSEFIALLP	SIRPRYYSIS	SSPRVDEKQA	900
SITVSVVSGE	AWSGYGEYKG	IASNYLAELQ	EGDTITCFIS	TPQSEFTLPK	DPETPLIMVG	960
PGTGVAPFRG	FVQARKQLKE	QGQSLGEAHL	YFGCRSPHED	YLYQEELENA	QSEGIITLHT	1020
AFSRMPNQPK	TYVQHVMEQD	GKKLIELLDK	GAHFYICGDG	SQMAPAVEAT	LMKSYADVHQ	1080
VSEADARLWL	QQLEEKGRYA	KDVA				1104

## SEQ ID NO:103

atgccaaagag	tgcctgaagt	cccaggtggt	ccattgttag	gaaatctggt	acaattgaag	60
gagaaaaagc	catacatgac	ttttacgaga	tgggcagcga	catatggacc	tatctatagt	120
atcaaaaactg	gggctacaag	tatggttggt	gtatcatcta	atgagatagc	caaggaggca	180
ttggtgacca	gattccaact	catatctaca	aggaacttat	ctaaagccct	gaaagtactt	240
acagcagata	agacaatggt	cgcaatgtca	gattatgatg	attatcataa	aacagttaag	300
agacacatac	tgaccgcctg	cttgggtcct	aatgcacaga	aaaagcatag	aattcacaga	360
gatatcatga	tggataacat	atctactcaa	cttcatgaat	tcgtgaaaaa	caaccagaa	420
caggaagagg	tagaccttag	aaaaatcttt	caatctgagt	tattcggtt	agctatgaga	480
caagccttag	gaaaggatgt	tgaagtttg	tacgttgaag	acctgaaat	cactatgaat	540
agagacgaaa	tctttcaagt	ccctgttggt	gatccaatga	tgggagcaat	cgatgttgat	600
tggagagact	tctttccata	cctaaagtgg	gtcccaaaca	aaaagttcga	aaatactatt	660
caacaaatgt	acatcagaag	agaagctggt	atgaaatctt	taatacaaga	gcacaaaaag	720
agaatagcgt	caggcgaaaa	gctaaaatgt	tatatcgatt	accttttatt	tgaagctcaa	780
actttaaccg	atcagcaact	attgatgtcc	ttgtgggaac	caatcattga	atcttcagat	840
acaacaatgg	tcacaacaga	atgggcaatg	tacgaattag	ctaaaaaccc	taaattgcaa	900
gataggttgt	acagagacat	taagtccgtc	tgtggatctg	aaaagataac	cgaagagcat	960
ctatcacagc	tgcctttacat	tacagctatt	ttccacgaaa	cactgagaag	acactcacca	1020
gttcctatca	ttcctctaat	acatgtacat	gaagataacc	ttctaggcgg	ctaccatggt	1080
ccttctggca	tcaccgttag	cgttaacatc	tacggttgca	acatggacaa	aaacgttttg	1140
gaaaatccag	aggaatggaa	cccagaaaag	ttcatgaaag	agaatgagac	aattgatttt	1200
caaaagacga	tggccttcgg	tggtggtaag	agagtttggt	ctggttcctt	gcaagccctt	1260
ttaaactgcat	ctattgggat	tgggagaatg	gttcaagagt	tcgaatggaa	actgaaggat	1320
atgactcaag	aggaagtga	cacgataggc	ctaactacac	aaatgttaag	accattgaga	1380
gctattatca	aacctaggat	cccatcaaga	ccaagtccta	gtaccgaaca	atctgcacaa	1440
aaagttagaa	aaaaagcaga	aaatgcacac	aatactccat	tgctagttct	ttatggttct	1500
aatatgggaa	cagcggaagg	aacggccagg	gatctagctg	acatagctat	gtccaaggga	1560
tttggcccg	aaatagcaac	cctggattcc	catgcaggta	acttgccaag	agaagggtgct	1620
gttctaatag	ttaccgctag	ctacaatggg	cacctccag	ataatgcgaa	gcagttcgtc	1680
gattgggttag	atcaagcatc	agcagatgaa	gttaagggtg	ttagatactc	tgtttttgga	1740
tgtggagata	agaattgggc	caccacatat	cagaagggtc	cggctttcat	cgatgaaatg	1800
cttgcgtgca	aaggggctga	aaatatagca	gatcgtggtg	aggccgacgc	aagcgacgat	1860
tttgagggta	cctatgagga	cacatgtggt	ctgatgttgc	cgcgtatttt		1920
aatctagaca	tagaaaattc	tgaagacaat	aaaagtcctt	tacttcttca	attcgtcgat	1980
agtgcgtg	acatgcctt	agcaaatgat	catggagcct	tttcaacgaa	cgtagtagcc	2040
agtaagggaac	ttcaacaacc	aggtagtgcc	agaagtacac	gtcacttgga	aattgaatta	2100
ccaaaagagg	catcctacca	agaaggtgac	catcttggtg	taatcccaag	aaactacgaa	2160
ggtatagtca	atagggtaac	ggcaagattt	gggctggatg	caagccaaca	gataagacta	2220
gaagcagaag	aagaaaaatt	ggcgacacct	ccactagcga	agacagtatc	cgttgaagaa	2280
ttattgcaat	acgtggaatt	gcaggatccc	gtcactagaa	cgcaattgag	agctatggca	2340
gcaaagactg	ttgttcacc	tcacaagggt	gaacttgaag	ctctacttga	aaaacaagca	2400
tacaaaagagc	aagtgtagc	accatgttag	aattgctgga	aaaatacccg		2460
gcatgcgaaa	tggaaattct	cgaatttctc	gcgttggtgc	caagtattcg	tcccagggtat	2520
tactcaattt	catcttcacc	aagggttgac	gagaaacagg	catctattac	cgtatctgtg	2580
gtctctggag	aagcttgga	tggttacgga	gaatacaagg	gtattgcttc	caattatctt	2640
gcagaactgc	aggaagggga	tacaattacc	tgccttattt	ctactcctca	atcagaattt	2700
actcttcgga	aggatccaga	aactccgcta	attatggtag	gtccgggaac	aggagtcgcc	2760
cctttcagag	gctttgtgca	agcaagggaag	caactaaaag	aacagggaca	aagtctgggt	2820
gaggcacatc	tatatctcgg	ttgcagatct	ccgcatgagg	attacttata	ccaagaagaa	2880

cttgaaaacg	cccaatcaga	aggtattatc	accttgcata	ctgcattcag	tagaatgcc	2940
aaccagccga	aaacttacgt	acagcatggt	atggagcaag	atggtaagaa	gttaattgag	3000
ctttttgata	agggcgccca	ctctacatt	tgccgacacg	gatcccaaat	ggcgccctgcc	3060
gttgaagcca	ccttgatgaa	atcatatgca	gatgttcac	aagtttcaga	agcggacgcc	3120
cgtctttggt	tacaacaact	agaggagaaa	ggaaggtatg	caaaagatgt	ttggttaa	3177

## SEQ ID NO:104

MPRVPEVPGV	PLLGNNLLQLK	EKKPYMTFTR	WAATYGPIS	IKTGATSMVV	VSSNEIAKEA	60
LVTRFQSIST	RNLSKALKVL	TADKTMVAMS	DYDDYHKTVK	RHILTAVLGP	NAQKKHRIHR	120
DIMMDNISTQ	LHEFVKNNPE	QEEVDLRKIF	QSELFGLAMR	QALGKDVESL	YVEDLKITMN	180
RDEIFQVLVV	DPMMAIDVD	WRDFFPYLKW	VPNKKFENTI	QQMYIRREAV	MKSLIKEHKK	240
RIASGEKLNS	YIDYLLSEAE	TLTDQQLLMS	LWEPIIESD	TTMVTTEWAM	YELAKNPKLQ	300
DRLYRDIKSV	CGSEKITEEH	LSQLPYITAI	FHETLRHSP	VPIPLRHVH	EDTVLGGYHV	360
PAGTELAVNI	YGCNMDKNVW	ENPEEWNPER	FMKENETIDF	QKTMAFGGGK	RVCAGSLQAL	420
LTASIGIGRM	VQEFEWKLD	MTQEEVNTIG	LTTQMLRPLR	AIKPRIPSR	PSPSTEQSAK	480
KVRKKAENAH	NTPLLVLVGS	NMGTAEGTAR	DLADIAMSKG	FAPQVATLDS	HAGNLPREGA	540
VLIVTASYNG	HPPDNAQOFV	DWLDQASADE	VKGVRSVFG	CGDKNWATTY	QKVPAFIDEM	600
LAAGKAENIA	DRGEADASDD	FEGTYEERE	HMWSDVAAYF	NLDIENSEDN	KSALLLQFVD	660
SAADMPLAKM	HGAFTSNVVA	SKELQPGSA	RSTRHLEIEL	PKEASYQEGD	HLGVIIPRNYE	720
GIVNRVTARF	GLDASQQIRL	EAEELKLAHL	PLAKTVSVEE	LLQYVELQDP	VRTQLRAMA	780
AKTVCPPHKV	ELEALLEKQA	YKEQVLAKRL	TMLELLEKYP	ACEMEFSEFI	ALLPSIRPRY	840
YSISSSPRVD	EKQASITVSV	VSGEAWSGYG	EYKGIASNYL	AELQEGDTIT	CFISTPQSEF	900
TLPKDPETPL	IMVGPGTGVA	PRGFVQARK	QLKEQGQSLG	EAHLVFGCRS	PHEDYLYQEE	960
LENAQSEGII	TLHTAFSRMP	NQPKTYVQHV	MEQDGKKLIE	LLDKGAHFYI	CGDGSQMAPA	1020
VEATLMKSYA	DVHQVSEADA	RLWLQLEEK	GRYAKDVW			1058

## SEQ ID NO:105

atgccaaagag	tgcctgaagt	cccagggtgtt	ccattgttag	gaaatctgtt	acaattgaag	60
gagaaaaagc	catacatgac	ttttacgaga	tgggcagcga	catatggacc	tatctatagt	120
atcaaaaactg	gggctacaag	tatgggtgtg	gtatcatcta	atgagatagc	caaggaggca	180
ttggtgacca	gattccaatc	catatctaca	aggaacttat	ctaaagccct	gaaagtactt	240
acagcagata	agacaatggt	cgcaatgtca	gatttatgat	attatcataa	aacagttaag	300
agacacatac	tgaccgcgct	cttgggtcct	aatgcacaga	aaaagcatag	aattcacaga	360
gatatcatga	tgataaacat	atctactcaa	cttcatgaat	tcgtgaaaaa	caaccagaa	420
caggaagagg	tagaccttag	aaaaatcttt	caatctgagt	tattcggctt	agctatgaga	480
caagcccttag	gaaaggatgt	tgaaagtgtg	tacgttgaag	acctgaaaaa	cactatgaat	540
agagacgaaa	tctttcaagt	ccttgttgtt	gatccaatga	tgaggagcaat	cgatgttgat	600
tgagagagact	tctttccata	cctaaaagtgg	gtcccaaaaca	aaaagttcga	aaatactatt	660
caacaaatgt	acatcagaag	agaagctgtt	atgaaatctt	taatcaaaga	gcacaaaaag	720
agaatagcgt	caggcgaaaa	gctaaatagt	tatatcgatt	accttttatc	tgaagctcaa	780
actttaaccg	atcagcaact	attgatgtcc	ttgtgggaac	caatcattga	atcttcagat	840
acaacaatgg	tcacaacaga	atgggcaatg	tacgaattag	ctaaaaacc	taaattgcaa	900
gatagggtgt	acagagacat	taagtccgtc	tgtggatctg	aaaagataac	cgaagagcat	960
ctatcacagc	tgacctacat	tacagctatt	ttccacgaaa	cactgagaag	acactcacca	1020
gttctctatca	ttcctctaag	acatgtacat	gaagataccg	ttctaggcgg	ctaccatgtt	1080
cctgtctggca	cagaactgtg	cgtaaacatc	tacggttgca	acatggacaa	aaacgttttg	1140
gaaaaatccag	aggaatggaa	cccagaaaaga	ttcatgaaag	agaatgagac	aattgatattt	1200
caaaaagacga	tgcccttcgg	tggtggtaag	agagtttgtg	ctggttcctt	gcaagccctt	1260
ttactgcat	ctattgggat	tgggagaaatg	gttcaagagt	tcgaatggaa	actgaaggat	1320
atgactcaag	aggaagtga	cacgataggc	ctaactacac	aatgtttaag	accattgaga	1380
gctattatca	aacctaggat	cccatcaaga	ccaagtccta	gtaccgaaca	atctgcaaaa	1440
aaagttagaa	aaaaagcaga	aaatgcacac	aatactccat	tgctagtctt	ttatggttct	1500
aatatgggaa	cagcggaagg	aacggccagg	gatctagctg	acatagctat	gtccaaggga	1560
tttgcctcgc	aagtagcaac	cctggattcc	catgcaggta	acttgccaag	agaaggtgct	1620
gttctaatag	ttaccgctag	ctacaatggg	caccctccag	ataatgcga	gcagttcgct	1680
gattgggttag	atcaagcatc	agcagatgaa	gttaagggtg	ttagatactc	tgtttttgga	1740
tgtggagata	agaattgggc	caccacatat	cagaagggtt	cggctttcat	cgatgaaatg	1800
cttgctgcaa	aaggggctga	aaatatagca	gatcgtgggt	aggccgacgc	aagcgacgat	1860
tttgagggta	cctatgagga	gtggagagag	cacatgtggt	ctgatgttgc	cgcgatattt	1920
aatctagaca	tagaaaattc	tgaagacaat	aaaagtgcct	tacttcttca	attcgtcgat	1980
agtgtctcgg	acatgccctt	agcaaaagatg	catggagcct	tttcaacgaa	cgtagtagcc	2040
agtaagggaac	ttcaacaacc	aggtagtgc	agaagtacac	gtcacttgga	aattgaatta	2100
ccaaaagagg	catcctacca	agaaggtgac	catcttggtg	taatcccaag	aaactacgaa	2160

gggtatagtca	ataggggtaac	ggcaagattt	gggctggatg	caagccaaca	gataagacta	2220
gaagcagaag	aagaaaaaatt	ggcgacacct	coactagcga	agacagtatc	cggtgaagaa	2280
ttattgcaat	acgtgggaatt	gcaggatccc	gtcactagaa	cgcaattgag	agctatggca	2340
gcaaagactg	tttgtccacc	tcacaagggt	gaacttgaag	ctctacttga	aaaacaagca	2400
tacaaagagc	aagtgtctagc	aaagagacta	accatgttag	aattgctgga	aaaatacccg	2460
gcatgcgaaa	tggaattctc	cgaatttata	gcgttgttgc	caagtattcg	tcccagggtat	2520
tactcaattt	catcttcacc	aagggttgac	gagaaacagg	catctattac	cgatatctgtg	2580
gtctctggag	aagcttggag	tggttacgga	gaatacaagg	gtattgcttc	caattatctt	2640
gcagaactgc	aggaagggga	tacaattacc	tgctttatct	ctactcctca	atcagaattt	2700
actcttccga	aggatccaga	aactccgtta	attatggtag	gtccgggaac	aggagtcgcc	2760
cctttcagag	gctttgtgca	agcaaggaag	caactaaaag	aacagggaca	aagtctgggt	2820
gaggcacatc	tatatttcgg	ttgcagatct	ccgcatgagg	attacttata	ccaagaagaa	2880
cttgaaaacg	cccaatcaga	aggtattatc	accttgcata	ctgcattcag	tagaatgcca	2940
aaccagccga	aaacttacgt	acagcatggt	atggagcaag	atggtaagaa	gttaattgag	3000
cttttgata	agggcgccca	cttctacatt	tgccgagcag	gatcccaaat	ggcgccgtgc	3060
gttggaagcca	ccttgatgaa	atcatacgca	gatgttcac	aagtttcaga	agcggacgcc	3120
cgtctttggt	tacaacaata	agaggagaaa	ggaaggtatg	caaaagatgt	tgcttaa	3177

## SEQ ID NO:106

MPRVPEVPGV	PLLGNLLQLK	EKKPYMTFTR	WAATYGPPIYS	IKTGATSMV	VSSNEIAKEA	60
LVTFRQSIST	RNLKALKVL	TADKTMVAMS	DYDDYHKT	RHILTAVLGP	NAQKKHRIHR	120
DIMDNIST	LHEFVKNPE	QEEVDLRKIF	QSEVGLAMR	QALGKDVESL	YVEDLKITMN	180
RDEIFQVLV	DPMMGAIDVD	WRDFFPYLKW	VPNKKFENTI	QQMYIRREAV	MKS LIKEHKK	240
RIASGEKINS	YIDYLLSEAG	TLTDQQLMS	LWEPIESSD	TTMVTTEWAM	YELAKNPKLQ	300
DRLYRDIKSV	CGSEKITEEH	LSQLPYITAI	FHETLRHSP	VPIIPLRHVH	EDTVLGGYHV	360
PAGTELAVNI	YGCNMDKNVW	ENPEEWNPER	FMKENETIDF	QKTMAFGGGK	RVCAGSLQAL	420
LTASIGIGRM	VQEFWKLLKD	MTQEEVNTIG	LTTQMLRPLR	AIKPRIPSR	PSPSTEQSAK	480
KVRKAENAH	NTPLLVLYGS	NMGTAEGTAR	DLADIAMSKG	FAPQVATLDS	HAGNLPREGA	540
VLIVTASYNG	HPPDNAKQFV	DWLDQASADE	VKGVRYSVFG	CGDKNWATTY	QKVPFIDEM	600
LAAGAENIA	DRGEADASDD	FEGTYEEWRE	HMWSDVAAYF	NLDIENSEDN	KSALLLQFVD	660
SAADMPLAKM	HGAFSTNVVA	SKELQQPGSA	RSTRLEIEL	PKEASYQEGD	HLGVI PRNVE	720
GIVNRVTARE	FAGSQQIRL	EAEELKLAHL	PLAKTVSVEE	LLQYVELQDP	VTRTQLRAMA	780
AKTVCPPHKV	ELEALLEKQA	YKEQVLAKRL	TMLELLEKYP	ACEMEFSEFI	ALLPSIRPRY	840
YSISSSPRVD	EKQASITVSV	VSGEAWSGYG	EYKGIASNYL	AELQEGDTIT	CFISTPQSEF	900
TLPKDPETPL	IMVGPGTGVA	PFRGFVQARK	QLKEQGQSLG	EAHLYFGCRS	PHEDYLYQEE	960
LENAQSEGII	TLHTAFSRMP	NQPKTYVQHV	MEQDGKKLIE	LLDKGAHFYI	CGDGSQMAPA	1020
VEATLMKSYA	DVHQVSEADA	RLWLQOLEEK	GRYAKDVA			1058

## SEQ ID NO:107

atggctacct	tggtggaaca	ttttcaagct	atgccattcg	ctattccaat	tgctttggct	60
gctttgtctt	ggttgttttt	gttctacatc	aaggtttctt	tcttctccaa	caaataccgt	120
caagctaaat	tgccaccagt	tccagttgtt	ccaggtttgc	cagttatttg	taatttgttg	180
caattgaaag	aaaagaagcc	ataccaaacc	tccactagat	gggctgaaga	atatggtcca	240
atctactcta	ttagaactgg	tgcttctact	atggttgtct	tgaacactac	tcaagttgcc	300
aaagaagcta	tggttaccag	atacttgtct	atctctacca	gaaagttgtc	caacgccttg	360
aaaattttga	ccgctgataa	gtgcattggt	gccattttctg	attacaacga	tttccacaag	420
atgatcaaga	gatatactct	gtctaacggt	ttgggtccat	ctgccccaaa	aagacataga	480
tctaacagag	ataccttgag	agccaacggt	tgctctagat	tgcatcccca	agttaagaac	540
tctccaagag	aagctgtcaa	ctttagaaga	gttttcgaat	gggaattatt	cggtatcgct	600
ttgaaacaag	ccttcggtta	ggatattgaa	aagccaatct	acgtcgaaga	attgggtact	660
actttgtcca	gagatgaaat	cttcaaggtt	ttggtcttgg	acattatgga	aggtgccatt	720
gaagttgatt	ggagagattt	tttccatac	ttgcgttggg	ttccaaacac	cagaatggaa	780
actaagatcc	aaagattata	ctttagaaga	aaggccgtta	tgaccgcctt	gattaacgaa	840
caaaagaaaa	gaattgcctc	cggtgaagaa	atcaactgct	acatcgattt	cctgttgaaa	900
gaaggtaaga	ccttgaccat	ggaccaaacc	tctatgttgt	tggtgggaaac	cgttattgaa	960
actgtcgata	ccacaatggt	tactactgaa	tggtctatgt	acgaagttgc	taaggattct	1020
aaaagacaag	acagattata	ccaagaaacc	caaaaggtct	gcggttctga	aatggttaca	1080
gaagaatact	tgtecccaat	gccatacttg	aatgctgttt	tccacgaaac	tttgagaaaa	1140
cattctccag	ctgcttttgt	tccattgaga	tatgctcatg	aagatactca	attgggttgt	1200
tattacattc	cagccggtac	tgaatttgcc	attaacatct	acggttgcaa	catggacaaa	1260
cccacatggg	aattctccga	agaatggaa	ttttggatcc	taagtttgac		1320
ccaatggact	tgtacaaaac	tatggctttt	gggtctggtg	aaagagtttg	cgctggttct	1380
ttacaagcta	tgttgattgc	ttgtccaacc	atcggttagat	tggttcaaga	atttgaatgg	1440

aagttgagag	atggtgaaga	agaaaacggt	gatactgttg	gtttgaccac	ccataagaga	1500
tatccaatgc	atgctatttt	gaagccaaga	tctccatcaa	gaccaagtcc	tagtaccgaa	1560
caatctgcaa	aaaaagttag	aaaaaaagca	gaaaatgcac	acaatactcc	attgctagtt	1620
ctttatggtt	ctaatatggg	aacagcggaa	ggaacggcca	gggatctagc	tgacatagct	1680
atgtccaagg	gatttgcccc	gcaagtagca	accctggatt	cccatgcagg	taacttgcca	1740
agagaagggtg	ctgttctaata	agttaccgct	agctacaatg	ggcaccctcc	agataatgcg	1800
aagcagttcg	tcgatttggtt	agatcaagca	tcagcagatg	aagttaaggg	tgtagatatac	1860
tctgtttttg	gatgtggaga	taagaattgg	gccaccacat	atcagaaggt	tccggctttc	1920
atcgatgaaa	tgcttgctgc	aaaaggggct	gaaaatatag	cagatcgtgg	tgaggccgac	1980
gcaagcgacg	attttgaggg	tacctatgag	gagtgaggag	agcacatgtg	gtctgatgtt	2040
gccgcgtatt	ttaatctaga	catagaaaat	tctgaagaca	ataaaagtgc	cttacttctt	2100
caattcgctg	atagtgcctg	ggacatgccc	ttagcaaaag	tgcatggagc	cttttcaacg	2160
aacgtagtag	ccagttaagg	acttcaacaa	ccaggtagtg	ccagaagtac	acgtcacttg	2220
gaaattgaat	taccaaaga	ggcatcctac	caagaagggtg	accatcttgg	tgtaatccca	2280
agaaactacg	aaggtatagt	caatagggta	acggcaagat	ttgggctgga	tgcaagccaa	2340
cagataagac	tagaagcaga	agaagaaaa	ttggcgcacc	ttccactagc	gaagacagta	2400
tccgttgaag	aattattgca	atacgtggaa	ttgcaggatc	ccgtcactag	aacgcaattg	2460
agagctatgg	cagcaaagac	tgtttgtcca	cctcacaagg	ttgaacttga	agctctactt	2520
gaaaaacaag	catacaaaag	gcaagtgcct	gcaaagagac	taaccatgtt	agaattgctg	2580
gaaaaatacc	cggcatgcga	aatggaaattc	tccgaattta	tcgcgttggt	gccaaagtatt	2640
cgtcccaggt	atgactcaat	ttcatcttca	ccaagggttg	acgagaaaca	ggcatctatt	2700
accgtatctg	tggtctctgg	agaagcttgg	agtggttacg	gagaatacaa	gggtattgct	2760
tccaattatc	ttgcagaact	gcaggaagg	gatacaatta	cctgctttat	ttctactcct	2820
caatcagaat	ttactcttcc	gaaggatcca	gaaactccgt	taattatggt	aggtccggga	2880
acaggagtgc	cccctttcag	aggctttgtg	caagcaagga	agcaactaaa	agaacaggga	2940
caaagtctgg	gtgaggcaca	tctatatattc	ggttgcatga	ctccgcatga	ggattactta	3000
taccaagaag	aacttgaaaa	cgcccaatca	gaagggtatta	tcaccttgca	tactgcattc	3060
agtagaatgc	caaaccagcc	gaaaacttac	gtacagcatg	ttatggagca	agatggtaag	3120
aagttaattg	agcttttgga	taaggcgcc	cacttctaca	tttgccggca	cggatcccaa	3180
atggcgctg	ccgttgaagc	caccttgatg	aaatcatatg	cagatgttca	tcaagtttca	3240
gaagcggacg	ccgctctttg	gttacaacaa	ctagaggaga	aaggaaggta	tgcaaaagat	3300
gtttggttaa						3309

## SEQ ID NO:108

MATLLEHFQA	MPFAIPIALA	ALSWLFLFYI	KVSFFSNKSA	QAKLPPVPV	PGLPVIGNLL	60
QLKEKKPYQT	FTRWAEYGP	IYSIRTGAST	MVVLNNTQVA	KEAMVTRYLS	ISTRKLSNAL	120
KILTADKCMV	AISDYNDHFK	MIKRYILSNV	LGPSAQKRHR	SNRDTLRANV	CSRLHSQVKN	180
SPREAVNFR	VFEWELFGIA	LKQAFGKDIE	KPIYVEELGT	TLRDEIFKV	LVL DIMEGAI	240
EVDWRDFFPY	LRWIPNTRME	TKIQRLYFRR	KAVMTALINE	QKKRIASGEE	INCYIDFLLK	300
EGKTLTMDQI	SMLLWETVIE	TADTMTVTE	WAMYEVAKDS	KRQDRLYQEI	QKVCSEMVT	360
EEYLSQLPYL	NAVFHETLRK	HSPAALVPLR	YAHEDTQLGG	YYIPAGTEIA	INIYGCNMDK	420
HQWESPEEWK	PERFLDPKFD	PMDLYKTMAF	GAGKRCVAGS	LQAMLIACPT	IGRLVQEFEW	480
KLRDGEENNV	DTVGLTTHKR	YPMHAILKPR	SPSRPSFSTE	QSARKVRKKA	ENAHNTPLL	540
LYGSNMGTAE	GTARDLADIA	MSKGFAPQVA	TLDSHAGNLP	REGAVLIVTA	SYNGHPPDNA	600
QKFVDWLDQA	SADEVKGVRY	SVFSGCDKNW	ATTYQKVPF	IDEMLAAKGA	ENIADRGEAD	660
ASDDFEGTYE	EWREHMWSDV	AAYFNLDIEN	SEDNKSALLL	QFVDSAADMP	LAKMHGAFST	720
NVVASKELOQ	PGSARSTRHL	EIELPKEASY	QEGDHLGVIP	RNYEGIVNRV	TARFGLDASQ	780
QIRLEAEEEK	LAHLPLAKTV	SVEELLQYVE	LQDPVTRTQL	RAMAAKTVC	PHKVELEALL	840
EKQAYKEQVL	AKRLTMLELL	EKYPACEMEF	SEFIALLP	RPRYYSISS	PRVDEKQASI	900
TVSVSGEAW	SGYGEYKGIA	SNYLAELQEG	DTITCFISTP	QSEFTLPKDP	ETPLIMVPG	960
TGVAPFRGFV	QARKQLKEQG	QSLGEAHLFY	GCRSPHEDYL	YQEELENAQS	EGIITLHTAF	1020
SRMPNQPKTY	VQHVMEQDQK	KLIELLDKGA	HFYICGDGSQ	MAPAVEATLM	KSYADVHQVS	1080
EADARLWLQ	LEEKGRYAKD	VW				1102

## SEQ ID NO:109

atggctacct	tggtggaaca	ttttcaagct	atgccattcg	ctattccaat	tgctttggct	60
gctttgtctt	ggttggtttt	gtttacatc	aaggtttctt	tcttctocaa	caaattccgct	120
caagctaaat	tgccaccagt	tccagttggt	ccaggtttgc	cagttattgg	taattttgtg	180
caattgaaag	aaaagaagcc	ataccaaacc	ttcactagat	gggctgaaga	atatggtcca	240
atctactcta	ttagaactgg	tgcttctact	atggttgctc	tgaacactac	tcaagttgcc	300
aaagaagcta	tggttaccga	atacttgtct	atctctacca	gaaagttgtc	caacgccttg	360
aaaattttga	ccgctgataa	gtgcatggtt	gccattttctg	attacaacga	tttccacaag	420
atgatcaaga	gatatatctt	gtctaacggt	ttgggtccat	ctgcccaaaa	aagacataga	480

tctaacagag	ataccttgag	agccaacggt	tgttctagat	tgcattccca	agttaagaac	540
tctccaagag	aagctgtcaa	ctttagaaga	gttttcgaat	gggaattatt	cggtatcgct	600
ttgaacaag	ccttcggtaa	ggatattgaa	aagccaatct	acgtcgaaga	attgggtact	660
actttgtcca	gagatgaaat	cttcaagggt	ttggtcttgg	acattatgga	agggtgccatt	720
gaagttgatt	ggagagattt	tttcccatat	ttgcgttgga	ttccaaacac	cagaatggaa	780
actaagatcc	aaagattata	ctttagaaga	aaggccggtta	tgaccgcctt	gattaacgaa	840
caaaagaaaa	gaattgcctc	cgggtgaagaa	atcaactgct	acatcgattt	cttggtgaaa	900
gaaggttaaga	ccttgaccat	ggaccaaaatc	tctatgttgt	tgtgggaaac	cgttattgaa	960
actgctgata	ccacaatggt	tactactgaa	tgggctatgt	acgaagttgc	taaggattct	1020
aaaagacaag	acagattata	ccaagaaatc	caaaagggtct	gcggttctga	aatgggtaca	1080
gaagaatact	tgtcccaatt	gccatacttg	aatgctgttt	tccacgaaac	tttgagaaaa	1140
cattctccag	ctgctttggt	tccatttgaga	tatgctcatg	aagatactca	attgggtggt	1200
tattacattc	cagccgggtac	tgaattggcc	attaacatct	acggttgcaa	catggacaaa	1260
caccaatggg	aatctccaga	agaattggaag	ccagaaagat	ttttggatcc	taagtttgac	1320
ccaatggact	tgtacaaaac	tatggctttt	ggtgctggtta	aaagagtttg	cgctggttct	1380
ttacaagcta	tgttgattgc	ttgtccaacc	atcggtagat	tggttcaaga	atttgaatgg	1440
aagttgagag	atggtgaaga	agaaaacggt	gatactgttg	gtttgaccac	ccataagaga	1500
tatccaatgc	atgctatttt	gaagccaaga	tctccatcaa	gaccaagtcc	tagtaccgaa	1560
caatctgcaa	aaaaagttag	aaaaaaagca	gaaaaatgcac	acaatactcc	attgctagtt	1620
ctttatggtt	ctaatatggg	aacagcggaa	ggaacggcca	gggatctagc	tgacatagct	1680
atgtccaagg	gatttgcccc	gcaagtagca	accctggatt	cccatgcagg	taacttgcca	1740
agagaagggtg	ctgttctaata	agttaccgct	agctacaatg	ggcaccctcc	agataatgctg	1800
aagcagttcg	tgcattgggt	agatcaagca	tcagcagatg	aagttaaggg	tgttagatac	1860
tctgtttttg	gatgtggaga	taagaattgg	gccaccacat	atcagaaggt	tccggctttc	1920
atcgatgaaa	tgcttgctgc	aaaaggggct	gaaaaatatag	cagatcggtg	tgaggccgac	1980
gcaagcgacg	attttgaggg	tacctatgag	gagtgagag	agcacatgtg	gtctgatgtt	2040
gccgcgtatt	ttaatctaga	catagaaaaat	tctgaagaca	ataaaagtgc	cttactcttt	2100
caattcgctg	atagtgctgc	ggacatgccc	ttagcaaaaga	tgcatggagc	cttttcaacg	2160
aacgtagtag	ccagtaagga	acttcaacaa	ccaggtagt	ccagaagtac	acgtcacttg	2220
gaaattgaat	taccaaaaaga	ggcatcctac	caagaagggtg	accatcttgg	tgtaatccca	2280
agaaactacg	aaggtatagt	caatagggtta	acggcaagat	ttgggctgga	tgcaagccaa	2340
cagataagac	tagaagcaga	agaagaaaaa	ttggcgccacc	ttccactagc	gaagacagta	2400
tccgttgaa	aattatttga	atacgtggaa	ttgcaggatc	ccgtcactag	aacgcaattg	2460
agagctatgg	cagcaaaagac	tgtttgtcca	cctcacaaag	ttgaacttga	agctctactt	2520
gaaaaacaag	catacaaaaga	gcaagtgtta	gcaaaagac	taaccatgtt	agaattgctg	2580
gaaaaatacc	cggcatgcga	aatggaattc	tccgaattta	tccggttgtt	gccaaagtatt	2640
cgtcccagggt	attactcaat	ttcatcttca	ccaagggttg	acgagaaaaca	ggcatctatt	2700
accgtatctg	tggctctctg	agaagcttgg	agtgtttacg	gagaatacaa	gggtattgct	2760
tccaattatc	ttgcagaact	gcaggaagg	gatacaatta	cctgctttat	ttctactcct	2820
caatcagaat	ttactcttcc	gaaggatcca	gaaactccgt	taattatggt	agggtccggga	2880
acagaggtcg	cccttttcag	aggctttgtg	caagcaagga	agcaactaaa	agaacaggga	2940
caaagtctgg	gtgaggcaca	tctatatattc	ggttgcatg	ctccgcatga	ggattactta	3000
taccaagaag	aacttgaaaa	cgcccaatca	gaaggtatta	tcaccttgca	tactgcattc	3060
agtagaatgc	caaaccagcc	gaaaacttac	gtacagcatg	ttatggagca	agatggtaag	3120
aagttaattg	agcttttggg	taagggcgcc	cacttctaca	tttgcggcga	cggatcccaa	3180
atggcgctcg	cggttgaagc	caacttgatg	aaatcatatg	cagatgttca	tcaagtttca	3240
gaagcggacg	ccgctctttg	gttacaacaa	ctagaggaga	aaggaaggta	tgcaaaagat	3300
gttgcttaa						3309

## SEQ ID NO:110

MATLLEHFQA	MPFAIPIALA	ALSWLFLFYI	KVSFFSNKSA	QAKLPVPV	PGLPVIGNLL	60
QLKEKKPYQT	FTRWAEYGP	IYSIRTGAST	MVVLNNTQVA	KEAMVTRYLS	ISTRKLSNAL	120
KILTADKCMV	AISDYNDFHK	MIKRYILSNV	LGPSAQKRHR	SNRDTLRANV	CSRLHSQVKN	180
SPREAVNFRR	VFEWELFGIA	LKQAFGKDIE	KPIYVEELGT	TLRDEIFKV	LVL DIMEGAI	240
EVDWRDFFPY	LRWIPNTRME	TKIQRLYFRR	KAVMTALINE	QKKRIASGEE	INCYIDFLK	300
EGKTLTMDQI	SMLLWETVIE	TADTTMTVTE	WAMYEVAKDS	KRQDRLYQEI	QKVCSEMVT	360
EEYLSQLPYL	NAVHETLRK	HSPAALVPLR	YAHEDTQLGG	YYIPAGTEIA	INIYGCNMDK	420
HQWESPEEWK	PERFLDPKFD	PMDLYKTMAF	GAGKRVCAGS	LQAMLIACPT	IGRLVQEFEW	480
KLRDGEENV	DTVGLTTHKR	YPMHAILKPR	SPSRPSPSTE	QSAKKVRKKA	ENAHNTPLL	540
LYGSNMGTAE	GTARDLADIA	MSKGFAPQVA	TLD SHAGNLP	REGAVLIVTA	SYNGHPPDNA	600
KQFVDWLDQA	SADDEVKGVRY	SVFGCGDKNW	ATYQKVPAP	IDEMLAAKGA	ENIADRGEAD	660
ASDDFEGTYE	EWREHMWSDV	AAFYNLDIEN	SEDNKSALLL	QFVDSAADMP	LAKMHGAFST	720
NVVASKELQQ	PGSARSTRHL	EIELPKESY	QEGDHLGVIP	RNYEGIVNRV	TARFGLDASQ	780
QIRLEAEEEE	LAHLPLAKTV	SVEELLQYVE	LQDPVTRTQL	RAMAAKTVC	PHKVELEALL	840
EKQAYKEQVL	AKRLTMELEL	EKYPACEMEF	SEPIALLPSI	RPRYYSISS	PRVDEKQASI	900

TVSVVSGEAW	SGYGEYKGIA	SNYLAELQEG	DTITCFISTP	QSEFTLPKDP	ETPLIMVGP	960
TGVAPFRGFV	QARKQLKEQG	QSLGEAHLYF	GCRSPHEDYL	YQEELNAQS	EGIITLHTAF	1020
SRMPNQPKTY	VQHVMQDQK	KLIELLDKGA	HFYICGDSQ	MAPAVEATLM	KSYADVHQVS	1080
EADARLWLQQ	LEEKGRYAKD	VA				1102

## SEQ ID NO:111

atggttccag	gtttgccagt	tattggtaat	ttgttgcaat	tgaagaaaa	gaagccatac	60
caaaccctca	ctagatgggc	tgaagaatat	ggtccaatct	actctattag	aactgggtgct	120
tctactatgg	ttgtcttgaa	cactactcaa	gttgccaaag	aagctatggt	taccagatac	180
ttgtctatct	ctaccagaaa	gttgtccaac	gccttgaaaa	ttttgaccgc	tgataagtgc	240
atggttgcca	tttctgatta	caacgatttc	cacaagatga	tcaagagata	tatcttgtct	300
aacgttttgg	gtccatctgc	ccaaaaaaga	catagatcta	acagagatac	cttgagagcc	360
aacgttttgt	ctagattgca	ttcccaagtt	aagaactctc	caagagaagc	tgtaactttt	420
agaagagttt	tcgaatggga	attattcggt	atcgctttga	aacaagcctt	cggttaaggat	480
attgaaaagc	caatctacgt	cgaagaattg	ggtactactt	tgccagaga	tgaaatcttc	540
aaggttttgg	tcttgacat	tatggaaggt	gccattgaag	ttgattggag	agattttttc	600
ccatacttgc	gttggtattcc	aaacaccaga	atggaaacta	agatccaaag	attatacttt	660
agaagaaagg	ccgttatgac	cgcttgatt	aacgaacaaa	agaaaagaat	tgctccgggt	720
gaagaaatca	actgctacat	cgatttcttg	ttgaaagaag	gtaagacctt	gaccatggac	780
caaatctcta	tggttgtgtg	ggaaaaccgtt	attgaaactg	ctgataccac	aatgggttact	840
actgaatggg	ctatgtacga	agttgctaag	gattctaaaa	gacaagacag	attataccaa	900
gaaatccaaa	aggtctgctg	ttctgaaatg	gttacagaag	aatacttgct	ccaattgcca	960
tacttgaaatg	ctgttttcca	cgaaaactttg	agaaaacatt	ctccagctgc	tttggttcca	1020
ttgagatatg	ctcatgaaga	tactcaattg	ggtggttatt	acattccagc	cggtactgaa	1080
attgcatata	acatctacgg	ttgcaacatg	gacaaacacc	aatgggaatc	tccagaagaa	1140
tggaagccag	aaagattttt	ggatcctaag	tttgacccaa	tggaacttga	caaaactatg	1200
gcttttggtg	ctggtaaaag	agtttgctgt	ggttctttac	aagctatggt	gattgcttgt	1260
ccaaccatcg	gtagattggg	tcaagaattt	gaatggaagt	tgagagatgg	tgaagaagaa	1320
aacgttgata	ctgttggttt	gaccacccat	aagagatatc	caatgcatgc	tattttgaag	1380
ccaagatctc	catcaagacc	aagtcctagt	accgaacaat	ctgcaaaaaa	agttagaaaa	1440
aaagcagaaa	atgcacacaa	tactccattg	ctagttcttt	atggttctaa	tatgggaaca	1500
gcggaaggaa	cgccacaggga	tctagctgac	atagctatgt	ccaagggtat	tgccccgcaa	1560
gtagcaaccc	tggaattccca	tgcaggtaac	ttgccaagag	aagggtgctgt	tctaatagtt	1620
accgctagct	acaatgggca	ccctccagat	aatgcgaagc	agttcgtcga	ttggttagat	1680
caagcatcag	cagatgaagt	taagggtgtt	agatactctg	tttttgatg	tgagataaag	1740
aattgggcca	ccacatatca	gaaggttocg	gctttcatcg	atgaaatgct	tgctgcaaaa	1800
gggctgaaa	atatagcaga	tcgtggtgag	gccgacgcaa	gcgacgattt	tgagggtacc	1860
tatagaggag	agagagagca	catgtgtctc	gatgttgccg	cgtattttaa	tctagacata	1920
gaaaattctg	aagacaataa	aagtgcctta	cttcttcaat	tcgtcgatag	tgctgaggac	1980
atgcccttag	caaagatgca	tggagccttt	tcaacgaacg	tagtagccag	taaggaactt	2040
caacaaccag	gtagtgccag	aagtacacgt	cacttggaag	ttgaattacc	aaaagaggca	2100
tcttaccagg	aaggtagcca	tcttggtgta	atcccaagaa	actacgaagg	tatagtcaat	2160
agggttaacg	caagattttg	gctggatgca	agccaacaga	taagactaga	agcagaagaa	2220
gaaaaatttg	cgacaccttc	actagcgaag	acagtatccg	ttgaagaatt	attgcaatac	2280
gtggaattgc	aggatcccg	cactagaacg	caattgagag	ctatggcagc	aaagactgtt	2340
tgtccacctc	acaaggttga	acttgaagct	ctacttgaaa	aacaagcata	caaagagcaa	2400
gtgctagcaa	agagactaac	catgttagaa	ttgctggaaa	aatacccgcc	atgcgaaatg	2460
gaattctccg	aatttatcgc	gttggttgcca	agtattcgct	ccaggatatta	ctcaatttca	2520
tcttacccaa	gggttgacga	gaaacaggca	tctattaccg	tatctgtggt	ctctggagaa	2580
gcttgagatg	gttacggaga	atacaagggt	attgcttcca	attatcttgc	agaactgcag	2640
gaaggggata	caattacctg	ctttatttct	actcctcaat	cagaatttac	tcttccgaag	2700
gatccagaaa	ctccgttaat	tatggttaggt	ccgggaacag	gagtcgcccc	tttcagaggc	2760
tttgtgcaag	caaggaagca	actaaaagaa	cagggaacaaa	gtctgggtga	ggcacatcta	2820
tatttcggtt	gcagatctcc	gcatgaggat	tactttatacc	aagaagaact	tgaaaacgcc	2880
caatcagaag	gtattatcac	cttgcatact	gcattcagta	gaatgccaaa	ccagccgaaa	2940
acttacgtac	agcatgttat	ggagcaagat	ggtaagaagt	taattgagct	tttgataaag	3000
ggcgccact	tctacatttg	cgcgacgga	tcccaaatgg	cgctgccgt	tgaagccacc	3060
ttgatgaaat	catatgcaga	tgttcatcaa	gtttcagaag	cggacgcccg	tctttggtta	3120
caacaactag	aggagaaagg	aaggtatgca	aaagatgttg	cttaa		3165

## SEQ ID NO:112

MVPGLPVIGN	LLQLKEKKPY	QTFTRWAEY	GPIYSIRTGA	STMVLTNTQ	VAKEAMVTRY	60
LSISTRKLSN	ALKILTADKC	MVAISDYND	HKMIKRYILS	NVLGPSAQKR	HRSNRDTLRA	120



NVCSRLHSQV	KNSPREAVNF	RRVFEWELFG	IALKQAFGKD	IEKPIYVEEL	GTTLSRDEIF	180
KVLVLDMIEG	AIEVDWRDFF	PYLRLWIPNTR	METKIQRLYF	RRKAVMTALI	NEQKKRIASG	240
EEINCYIDFL	LKEGKTLTMD	QISMLLWETV	IETADTMVT	TEWAMYEVAK	DSKRQDRLYQ	300
EIQKVCGSEM	VTEEYLSQLP	YLNNAVHETL	RKHSPAALVP	LRyahEDTQL	GGYYPAGTE	360
IAINIYGCNM	DKHQWESPEE	WKPERFLDPK	FDPMDLYKTM	AFGAGKRVCA	GSLQAMLIAC	420
PTIGRLVQEF	EWKLRDGEEE	NVDTVGLTTH	KRYPMHAILK	PRSPSRPSPS	TEQSAKKVRK	480
KAENAHNTPL	LVLYGSNMGT	AEGTARDLAD	IAMSKGFAPQ	VATLDSHAGN	LPREGAVLIV	540
TASYNHPPD	NAKQFVDWLD	QASADEVKGV	RYSVFGCGDK	NWATTYQKVP	AFIDEMLAAC	600
GAENIADRGE	ADASDDFEGT	YEWEHREHWS	DVAAYFNLDI	ENSEDNKSAL	LLQFVDSAAD	660
MPLAKMHGAF	STNVVASKEL	QQPGSARSTR	HLEIELPKEA	SYQEGDHLGV	IPRNYEGIVN	720
RVTARFGLDA	SQQIRLEAEE	EKLAHLPLAK	TVSVEELLQY	VELQDPVTRT	QLRAMAAKTV	780
CPPHKVELEA	LLEKQAYKEQ	VLAKRLTMLE	LLEKYFACEM	EFSEFIALLP	SIRPRYYSIS	840
SSPRVDEKQA	SITVSUVSGE	AWSGYGEYKG	IASNYLAELQ	EGDTITCFIS	TPQSEFTLPK	900
DPETELIMVG	PGTGVAFFRG	FVQARKQLKE	QGQSLGEAHL	YFGCRSPHED	YLYQEELENA	960
QSEGIITLHT	AFSRMPNQPK	TYVQHVMEQD	GKKLIELLDD	GAHFYICGDG	SQMAPAVEAT	1020
LMKSYADVHQ	VSEADARLWL	QQLEEKGRYA	KDVA			1054

## SEQ ID NO:113

atgaccagtt	tgtccaaaag	cttcatgcag	agtggacgaa	tctgcgcagc	atgtttctat	60
ctgttattca	cactactttc	aattccaatc	tcgtttaaag	ttgggtggtt	ggaatgcggg	120
ctttccttca	cgttgacact	gttcaacttta	tatttcataa	ctacgactct	taacgtgttg	180
gcaagacgac	atggaggaag	actatacatt	ttttttacca	actgtctgta	ttactcacia	240
ctttttatca	ttgcatcttt	gctataacctg	tttttgcctg	gattttctaa	tgatgagttg	300
ggaaacgttc	tgaaaaataa	atataatgag	tcggagtcgt	tcctggaagc	tttgaaaaat	360
agcttgaatt	ccaatcaaat	taactacgtc	ttatatattt	actactatcg	atttgtttgta	420
caaccgtggc	aattcgtgct	taccaagtc	acaccttttt	ttactctatc	ggaaggtttt	480
ttcactattt	tagccattca	ggccgctcgg	gaaactaata	gatggttatc	aatgactttg	540
aattcaaaaca	cgtggattat	ttcctcattg	ttaacctccg	gaggtgtgat	taccgcatcg	600
ctgtactatt	tgtatcggat	ttatgtcacc	cccatatggc	cgttatccat	ccaaacggcg	660
tccttattag	gacttggttt	gtctatggta	tgtggactgg	ggttgatagg	tattgtgagt	720
caaaaaggat	ccgtcataga	gagctcttta	ttttttgcgt	atattgttcg	ttgtatttat	780
gaaattttcc	ccaaatttagc	tactaccgcg	actgatgaaa	ttttaaattt	gttcaaagac	840
gtctggcaga	aacatcaaag	aaatctgccc	acagctgaca	atcttttgtg	ctactttcat	900
aatgtcatat	tgaaaaatgc	agaggtgtta	tgggggtcct	ttatccctag	aggaagaaaag	960
aaaaccgggtg	attttcatga	taaaactcatt	agcattctat	cattcgaaaa	agtatccttg	1020
atatctaaac	cattttggaa	atttttcaag	aatttcacct	ttagtggttc	gctatccatt	1080
aatgaatttt	tcaagtttac	aattaagatg	gcaagcgaat	cagtttcccc	agctatagta	1140
atcaatttat	gcttttagagt	tctgatgttt	tactcggcaa	cgaggattat	tccagcatta	1200
caaaagaaaa	atgacaaaaca	gttgcgcaag	agtcgcagga	tcatgaaggg	attgtatttg	1260
tacagtcctt	gcatattaat	tgctatgtat	actcacctga	ttttacaata	ttcaggtgag	1320
ctaaagaaaag	acctgtgcac	atgggggttg	agtgaaaagt	ggtttgccgt	agatcaacca	1380
gaaattatag	tagattcatg	gggaattttg	aactgggtgca	acattttctg	tactattttg	1440
gtatacgcta	cagaattaat	aggttctcgt	agttga			1476

## SEQ ID NO:114

MTSLSKSFMQ	SGRICAACFY	LLFTLLSIPI	SFKVGGLECG	LSFTVTLFTL	YFITTTLNVL	60
ARRHGGRLYI	FETSCLYYSQ	HFIASLLYL	FLSGFSNDEL	GNVLKNKYNE	SESFLEALKN	120
SLNSNQINYV	LYYYYYREVV	QPWFVLTKS	TPFFTLSEGF	FTILAIQAVG	ETNRWLSNDL	180
NSNTWIISSL	LTSGGVITAS	LYLYRIYVT	PIWPLSIQTA	SLLGFLSMV	CGLGLYIVS	240
QKGSVIESSL	FFAYIVRCIY	EISPKLATTA	TDEILNLFKD	VWQKHQRNLP	TADNLLCYFH	300
NVILKNAEVL	WGSFIPRGRK	KTGDFHDKLI	SILSFEKVS	ISKPFWKFFK	NFTFSVPLSI	360
NEFCQVTIKM	ASESVSPAIV	INLCFRVLMF	YSATRIIPAL	QRKNDKQLRK	SRRIMKGLYW	420
YSPCILIAMY	THLILQYSGE	LKKDLCIWGC	SEKWFGVDQP	EIIVDSWGF	NWCNIFCTIL	480
VYATELIGSG	S					491

## SEQ ID NO:115

agatctttat	gaagacatag	ctgcagaaga	aaaagcaaga	gctacatata	aatgggttaat	60
tgatatatca	gatgatcccg	atttaaacga	cagcttacga	tttttacgag	aaagagagat	120
tgttcactca	cagcgggttc	gcgaggccgt	ggagatttta	aaagatgaca	gagacaggaa	180
gaaaatcttt	taactagtaa	aaaaacatcc	cccttgccga	atgcaaacga	aaggagggat	240
gttttttggt	gtgactgcgt	tgattatgcg	ctagaactgc	agtgacaaga	aacaaccttt	300
aatttccctt	caacatcttt	ccaaactcgc	gtataactgt	attcacctcc	aatagattca	360

ccggttgcca	gtgccccatt	taacgctact	tttghtaacgg	taacggcaag	ttcttgaaac	420
agttttaactt	cttggttccaa	cactttccatg	cccgcgtatat	caagactttt	tgaacgatga	480
acattttatat	cttcttcttt	tgacaaccat	tgcccaaggt	gattcacaaa	aataagctca	540
tctgaaagta	attcttctaa	tagctctatg	ttattagaaa	gcatggctga	gcgaagcatt	600
tcttcgtatt	ctataactct	tgcttgattc	atttttaatc	ctcctttacg	ccttggttaa	660
ctcttttctta	tttccacggt	gcttttctct	taaacttctt	tcattaataa	ttcgtgctaa	720
attatgttaa	tagaggggat	aagtggaacta	attttctgta	agcactaaat	attctgaaat	780
actctgttaa	ttacctttaa	atggtataaa	attagaatga	aagaaccttt	tctttccact	840
tttctagtta	tctttttact	gaattatgcc	agttttttat	acttgtaatt	gtagcggaat	900
gaacgttcat	tccgtttttg	aaaagaggtg	ataaagtga	atctactcca	acaaaacaaa	960
aagcgatttt	ttctgcttcg	cttctgctgt	ttgcagaaa	aggggttgat	gcaaccacga	1020
tgccaatgat	tgacagaaat	gccaaagtag	gagcaggaac	aatttatcgc	tacttttaaaa	1080
ataaagaaa	ccttgtaaat	gaattatgcc	aacagcacgt	aaacgagttt	ttacagtga	1140
ttgaaagcgg	tctggcaaac	gagagagatg	gataccgaga	tggtgttcat	catatctttg	1200
aaggtatggt	gacatttact	aaaaaccatc	ctcgtgctct	tggttttatt	aaaactcata	1260
gccaaaggaac	ttttttaaca	gaagagagcc	gcttagcata	tcaaaagctg	gtggaatttg	1320
tttghtacgtt	cttcagagaa	ggacaaaagc	aaggtgtgat	tagaaatctt	cctgaaaaatg	1380
cgctaattgc	tatttttatt	ggaagtttca	tggaagtata	tgaaatgatt	gaaaatgact	1440
acttatcttt	aactgatgaa	cttcttaccg	gtgtagaaga	gagtctgtgg	gcagcactta	1500
gcagacaatc	atgaaaactta	acaagtgaag	gagggataac	atgacaatta	aagaaatgcc	1560
tcagccaaaa	acgtttggag	agcttaaaaa	tttaccgtta	ttaaacacag	ataaacccgtt	1620
tcaagctttg	atgaaaaattg	cggatgaatt	aggagaatc	tttaaatcgg	aggcgctgg	1680
tcgtgtaacg	cgctacttat	caagtcagcg	tctaattaaa	gaagcatgcg	atgaatcacg	1740
ctttgataaa	aacttaagtc	aagcgcttaa	atttgtacgt	gattttgcag	gagacgggtt	1800
atttacaagc	tggaacgcgt	aaaaaaattg	gaaaaaagcg	cataatatct	tacttccaag	1860
cttcagtcag	caggcaatga	aaggctatca	tgcatgatg	gtcgatatcg	cogtgcagct	1920
tgttcaaaag	tggaacgcgt	taaatgcaga	tgagcatatt	gaagtaccgg	aagacatgac	1980
acgtttaacg	cttgatacaa	ttggtctttg	cggtcttaac	tatcgcttta	acagctttta	2040
ccgagatcag	cctcatccat	ttattacaag	tatggtccgt	gcactggatg	aagcaatgaa	2100
caagctgcag	cgagcaaatc	cagacgaccc	agcttatgat	gaaaaacaagc	gccagtttca	2160
agaagatatc	aaggtgatga	acgacctagt	agataaaatt	attgcagatc	gcaaagcaag	2220
cgggtaacaa	agcgatgatt	tattaaacga	tatgctaaac	ggaaaaagatc	cagaaacggg	2280
tgagccgctt	gatgacgaga	acattcgcta	tcaaatattt	acattcttaa	ttgcgggaca	2340
cgaacaacaa	agtggtcttt	tatcatattg	gctgtatttc	ttagtgaata	atccacatgt	2400
attacaaaaa	gcagcagaag	aagcagcacg	agttctagta	gatcctgttc	caagctacaa	2460
acaagtcaaa	cagcttaaat	atgtcggcat	ggtcttaaac	gaagcgctgc	gcttatggcc	2520
aactgctcct	gogttttccc	tatatgcaaa	agaagatacg	gtgcttgagg	gagaatatcc	2580
tttagaaaaa	ggcgacgaac	taatggttct	gattcctcag	cttcaccgtg	ataaaaacat	2640
ttggggagac	gatgtggaag	agttccgtcc	agagcgtttt	gaaaaatccaa	gtgcgattcc	2700
gcagcatgct	ttttaaaccgt	ttggaacgg	tcagcgctgc	tgatcgggtc	agcagttcgc	2760
tcttcattgaa	gcaacgctgg	tacttggtat	gatgctaaaa	cactttgact	ttgaagatca	2820
tacaaactac	gagctggata	ttaaagaaac	tttaacgtta	aaacctgaag	gctttgtggt	2880
aaaagcaaaa	tcgaaaaaaa	ttccgcttgg	cggtattcct	tcacctagca	ctgaacagtc	2940
tgctaaaaaa	gtacgcaaaa	aggcagaaaa	cgctcataat	acgcccgtgc	ttgtgctata	3000
cggttcaaat	atgggaacag	ctgaaggaac	ggcgctgat	ttagcagata	ttgcaatgag	3060
caaaggattt	gcaccgcagg	tcgcaacgct	tgattcacac	gccggaatc	ttccgcgcga	3120
aggagctgta	tttaattgtaa	cggtctctta	taacggtcat	ccgcctgata	acgcaaaagca	3180
atttgtcgac	tggttagacc	aagcgtctgc	tgatgaagta	aaaggcgctc	gctactccgt	3240
atttgatgc	ggcgataaaa	actgggtac	tacgtatcaa	aaagtgcctg	cttttatcga	3300
tgaaacgctt	gccgctaaag	ggcgagaaaa	catcgctgac	cgcggtgaag	cagatgcaag	3360
cgacgacttt	gaaggcacat	atgaagaatg	gcgtgaacat	atgtggagtg	acgtagcagc	3420
ctactttaac	ctcgacattg	aaaacagtga	agataataaa	tctactcttt	cacttcaatt	3480
tgctgacagc	gccgcggata	tgccgcttgc	gaaaatgcac	ggtgcgtttt	caacgaacgt	3540
cgtagcaagc	aaagaacttc	aacagccagg	cagtgcacga	agcacgcgac	atcttgaaat	3600
tgaacttcca	aaagaagctt	cttatcaaga	aggagatcat	ttaggtgtta	ttcctcgcaa	3660
ctatgaagga	atagtaaac	gtgtaacagc	aaggttcggc	ctagatgcac	cacagcaaat	3720
ccgtctggaa	gcagaagaag	aaaaattagc	tcatttgcca	ctcgtataaa	cagtatccgt	3780
agaagagctt	ctgcaataacg	tggagcttca	agatcctggt	acgcgcacgc	agcttcgcgc	3840
aatggctgct	aaaaacggtc	gccgcgcga	taaatgagag	cttgaaagcct	tgcttgaaaa	3900
gcaagcctac	aaagaacaag	tgctggcaaa	acgtttaaca	atgcttgaac	tgcttgaaaa	3960
ataccgcggc	tgtgaaatga	aattcagcga	atttatcgcc	cttctgccaa	gcatacgccc	4020
cgctatttac	tcgattttct	catcacctcg	tgctgatgaa	aaacaagcaa	gcatacaggt	4080
cagcgttgct	tcaggagaa	cggtggagcg	atatggagaa	tataaaggaa	ttgcgtcgaa	4140
ctatcttgcc	gagctgcaag	aaggagatac	gattacgtgc	tttatttcca	caccgcagtc	4200
agaattttacg	ctgcaaaaag	accctgaaac	gccgcttatc	atggtcggac	cggaacagag	4260
cgctcgccgc	tttagaggct	ttgtgcaggc	gcgcaaacag	ctaaaagaac	aaggacagtc	4320
acttgagaa	gcacatttat	acttcggctg	ccgttcacct	catgaagact	atctgtatca	4380

agaagagctt	gaaaacgccc	aaagcgaagg	catcattacg	cttcataccg	ctttttctcg	4440
catgccaaat	cagccgaaaa	catacgttca	gcacgtaatg	gaacaagacg	gcaagaaatt	4500
gattgaactt	cttgatcaag	gagcgcactt	ctatatattg	ggagacggaa	gccaaatggc	4560
acctgccgtt	gaagcaacgc	ttatgaaaag	ctatgctgac	gttcaccaag	tgagtgaagc	4620
agacgctcgc	ttatggctgc	agcagctaga	agaaaaaggc	cgatacgcaa	aagacgtgtg	4680
ggctgggtaa	attaaaaaga	ggctaggata	aaagtagttt	agttggttga	aggaagatcc	4740
gaacgatgaa	tcgttcggat	ctttttattg	gtagagtaaa	cgtagatttc	atctatttag	4800
tgacttgtag	cggttgattg	gagggcaagg	tgaagactcc	aatcaaccgc	ggtgtcacat	4860
gcaagccata	cgaattccat	ttctcccatt	tattcgtctt	ttgtccccc	ttaattttta	4920
tagcgcctta	acgtttcttc	tgcgtagacg	cagatctt			4957

## SEQ ID NO:116

MTIKEMPQPK	TFGELKNLPL	LNTDKPVQAL	MKIADELGEI	FKFEAPGRVT	RYLSSQRLIK	60
EACDESFRDK	NLSQALKFVR	DFAGDGLFTS	WTHEKNWKA	HNILLPSFSQ	QAMKGYHAMM	120
VDIAVQLVGK	WERLNADEHI	EVPEMDTRLT	LDITIGLCGFN	YRFNSFYRDQ	PHPFITSMVR	180
ALDEAMNKLQ	RANPDDPAYD	ENKRQFQEDI	KVMNDLVDKI	IADRKASGEQ	SDDLTHMLN	240
GKDPETGEPL	DDENIRYQII	TFLIAGHETT	SGLLSFALYF	LVKNPHVLQK	AAEEAARVLV	300
DPVPSYKQVK	QLKYVGMVLN	EALRLWPTAP	AFSLYAKEDT	VLGGEYPLEK	GDELMVLIPQ	360
LHRDKTIWGD	WVEEFRPERL	ENPSAIPQHA	FKPFGNGQRA	CIGQQFALHE	ATLVLGMLK	420
HDFEDHTNY	ELDIKETLTL	KPEGFVVKAK	SKKIPLGGIP	SPSTEQSAKK	VRKKAENAHN	480
TPLLVLYGSN	MGTAEGTARD	LADIAMSKGF	APQVATLDSH	AGNLPREGAV	LIVTASYNGH	540
PPDNAKQFVD	WLDQASADEV	KGVRYSVFGC	GDKNWATTYQ	KVPAFIDETL	AAKGAENIAD	600
RGEADASDDF	EGTYEEWREH	MWSDVAAYFN	LDIENSEDNK	STLSLQFVDS	AADMPLAKMH	660
GAFSTNVVAS	KELQQPGSAR	STRHLEIELP	KEASYQEGDH	LGVI PRNYEG	IVNRVTARFG	720
LDASQQIRLE	ABEEKLAHLP	LAKTVSVEEL	LQYVELQDPV	TRTQLRAMAA	KTVCPPHKVE	780
LEALLEKQAY	KEQVLAKRLT	MLELLEKYP	CEMKSEFIA	LLPSIRPRY	SISSSPRVDE	840
KQASITVSVV	SGEAWSGYGE	YKGIASNYLA	ELQEGDTITC	FISTPQSEFT	LPKDPETPLI	900
MVGPGTGVP	FRGFVQARKQ	LKEQGQSLGE	AHLYFGCRSP	HEDYLYQEEL	ENAQSEGIIT	960
LHTAFSRMPN	QPKTYVQHVM	EQDGKKLIEL	LDQGAHFYIC	GDGSQMAPAV	EATLMKSYAD	1020
VHQVSEADAR	LWLQQLEEK	RYAKDVWAG				1049

## SEQ ID NO:117

ccaagtcccta	gtaccgaaca	atctgcaaaa	aaagtagtaa	aaaaagcaga	aaatgcacac	60
aatactccat	tgctagtctt	ttatggttct	aatatgggaa	cagcgggaag	aacggccagg	120
gatctagctg	acatagctat	gtccaaggga	tttgccccgc	aagtagcaac	cctggattcc	180
catgcaggta	acttgccaag	agaaggtgct	gttctaatag	ttaccgctag	ctacaatggg	240
caccctccag	ataatgcgaa	gcagttcgtc	gattggttag	atcaagcatc	agcagatgaa	300
gttaagggtg	ttagatactc	tgtttttgga	tgtggagata	agaattgggc	caccacatat	360
cagaaggttc	cggctttcat	cgatgaaatg	cttgctgcaa	aaggggctga	aaatatagca	420
gatcgtgggt	aggccgacgc	aagcgacgat	tttgagggtg	cctatgagga	gtggagagag	480
cacatgtggt	ctgatgttgc	cgcgattttt	aatctagaca	tagaaaattc	tgaagacaat	540
aaaagtgcc	tacttcttca	attcgtcgat	agtgtctcgg	acatgccctt	agcaaagatg	600
catggagcct	tttcaacgaa	cgtagtagcc	agtaaggaa	ttcaacaacc	aggtagtgcc	660
agaagtaac	gtcacttgga	aattgaatta	ccaaaagagg	catcctacca	agaagggtgac	720
catcttggtg	taatcccaag	aaactacgaa	ggtatagtca	atagggtaac	ggcaagattt	780
gggctggatg	caagccaaca	gataagacta	gaagcagaag	aagaaaaatt	ggcgacacct	840
ccactagcga	agacagtatc	cgttgaagaa	ttattgcaat	acgtggaatt	gcaggatccc	900
gtcactagaa	cgcaattgga	agctatggca	gcaaagactg	tttgtccacc	tcacaagggt	960
gaacttgaga	ctctacttga	aaaacaagca	tacaaagagc	aagtgtctagc	aaagagacta	1020
accatgttag	aattgtctga	aaaatacccg	gcatgcgaaa	tggaaattctc	cgaatttatc	1080
gcggtgtgtg	caagtattcg	ttcccaggat	tactcaattt	catcttcacc	aagggttgac	1140
gagaaacagg	catctattac	cgtatctgtg	gtctctggag	aagcttggag	tggttacgga	1200
gaatacaagg	gatctgttgc	caattatctt	gcagaactgc	aggaagggga	tacaattacc	1260
tgcttttatt	ctactcctca	atcagaattt	actcttccga	aggatccaga	aactccgtta	1320
attatggtag	gtccgggaac	aggagtcgcc	cctttcagag	gctttgtgca	agcaagggaag	1380
caactaaaag	aacagggaca	aagtctgggt	gaggcacatc	tatatttcgg	ttgcagatct	1440
ccgcatgagg	attacttata	ccaagaagaa	cttgaaaaacg	cccaatcaga	aggtattatc	1500
accttgcata	ctgcattcag	tagaatgcca	aaccagccga	aaacttacgt	acagcatgtt	1560
atggagcaag	atggtaagaa	gttaattgag	cttttgata	agggcgccca	cttctacatt	1620
tgccggcgacg	gatcccaaat	ggcgccgtcc	gttgaagcca	ccttgatgaa	atcatatgca	1680
gatgttcatc	aagtttcaga	agcggacgcc	cgtctttggt	tacaacaact	agaggagaaa	1740
ggaaggtatg	caaaagatgt	ttggttaa				1767

## SEQ ID NO:118

PSPSTEQSAK	KVRKKAENAH	NTPLLVLVYGS	NMGTAEGTAR	DLADIAMSKG	FAPQVATLDS	60
HAGNLPREGA	VLIIVTASYNG	HPPDNAKQFV	DWLDQASADE	VKGVRYSVFG	CGDKNWATTY	120
QKVPAFIDEM	LAAKGAENIA	DRGEADASDD	FEGTYEEWRE	HMWSDVAAYF	NLDIENSEDN	180
KSALLLQFVD	SAADMPLAKM	HGAFSTNVVA	SKELQQPGSA	RSTRHLEIEL	PKEASYQEGD	240
HLGVI PRNVE	GIVNRVTARF	GLDASQQIRL	EAEEEKLAHL	PLAKTVSVEE	LLQYVELQDP	300
VTRTQLRAMA	AKTVCPPHKV	ELEALLEKQA	YKEQVLAKRL	TMLELLEKYP	ACEMEFSEFI	360
ALLPSIRPRY	YSISSSPRVD	EKQASITVSV	VSGEAWSGYG	EYKGIASNYL	AELQEGDTIT	420
CFISTPQSEF	TLPKDPETPL	IMVGP GTGVA	PFRGFVQARK	QLKEQGQSLG	EAHLYFGCRS	480
PHEDYLYQEE	LENAQSEGII	TLHTAFSRMP	NQPKTYVQHV	MEQDGKKLIE	LLDKGAHFYI	540
CGDGSQMAPA	VEATLMKSYA	DVHQVSEADA	RLWLQQLLEK	GRYAKDVW		588

## SEQ ID NO:119

ccaagtccta	gtaccgaaca	atctgcaaaa	aaagttagaa	aaaaagcaga	aaatgcacac	60
aatactccat	tgctagttct	ttatggttct	aatatgggaa	cagcgggaagg	aacggccagg	120
gatctagctg	acatagctat	gtccaaggga	tttgccccgc	aagtagcaac	cctggattcc	180
catgcaggta	acttgccaag	agaagtgctg	gttccaatag	ttaccgctag	ctacaatggg	240
cacctccag	ataatgcgaa	gcagttcgct	gattggttag	atcaagcatc	agcagatgaa	300
gttaagggtg	ttagatactc	tggttttgga	tggtggagata	agaattgggc	caccacatat	360
cagaagggtc	cggcctttcat	cgatgaaatg	cttgctgcaa	aaggggctga	aaatatagca	420
gatcgtggtg	aggccgacgc	aagcgacgat	tttgagggtta	cctatgagga	gtggagagag	480
cacatgtggt	ctgatgttgc	cgcgatattt	aatctagaca	tagaaaaattc	tgaagacaat	540
aaaagtgctt	tacttcttca	attcgtcgat	agtgcgtcgg	acatgccctt	agcaaagatg	600
catggagcct	tttcaacgaa	cgtagtagcc	agtaagggaac	ttcaacaacc	aggtagtgcc	660
agaagtacac	gtcacttgga	aattgaatta	ccaaaagagg	catcctacca	agaaggtgac	720
catcttggtg	taatcccaag	aaactacgaa	ggtatagtca	atagggtaac	ggcaagattt	780
gggctggatg	caagccaaca	gataagacta	gaagcagaag	aagaaaaatt	ggcgcacctt	840
ccactagcga	agacagtatc	cgttgaagaa	ttattgcaat	acgtgggaatt	gcaggatccc	900
gtcactagaa	cgcaattgag	agctatggca	gcaaagactg	tttgtccacc	tcacaagggt	960
gaacttggaag	ctctacttga	aaaacaagca	tacaaagagc	aagtgcctagc	aaagagacta	1020
accatggttag	aattgctgga	aaaataccccg	gcatgcgaaa	tggaattctc	cgaatttatc	1080
gcggttggtg	caagtattcg	tcccagggtat	tactcaattt	catcttcacc	aaggggttgac	1140
gagaaacagg	catctattac	cgtatctgtg	gtctctggag	aagcttggag	tggttacgga	1200
gaatacaagg	gtattgcttc	caattatctt	gcagaactgc	aggaagggga	tacaattacc	1260
tgctttattt	ctactcctca	atcagaattt	actcttcgga	aggatccaga	aactccgtta	1320
attatggttag	gtccgggaac	aggagtcgcc	ccttttcagag	gctttgtgca	agcaagggaag	1380
caactaaaag	aacagggaca	aagtctgggt	gaggcacatc	tatatctcgg	ttgcagatct	1440
ccgcatgagg	attactttata	ccaagaagaa	cttgaaaacg	cccaatcaga	aggtattatc	1500
accttgcata	ctgcattcag	tagaatgccca	aaccagccga	aaacttacgt	acagcatggt	1560
atggagcaag	atggtaagaa	gttaattgag	cttttggtgata	agggcgccca	cttctacatt	1620
tgccgacgac	gatcccaaat	ggcgctgcc	gttgaaagcca	ccttgatgaa	atcatatgca	1680
gatgttcac	aagtttcaga	agcggacgcc	cgtctttggt	tacaacaact	agaggagaaa	1740
ggaaggtag	caaaagatgt	tgcttaa				1767

## SEQ ID NO:120

PSPSTEQSAK	KVRKKAENAH	NTPLLVLVYGS	NMGTAEGTAR	DLADIAMSKG	FAPQVATLDS	60
HAGNLPREGA	VLIIVTASYNG	HPPDNAKQFV	DWLDQASADE	VKGVRYSVFG	CGDKNWATTY	120
QKVPAFIDEM	LAAKGAENIA	DRGEADASDD	FEGTYEEWRE	HMWSDVAAYF	NLDIENSEDN	180
KSALLLQFVD	SAADMPLAKM	HGAFSTNVVA	SKELQQPGSA	RSTRHLEIEL	PKEASYQEGD	240
HLGVI PRNVE	GIVNRVTARF	GLDASQQIRL	EAEEEKLAHL	PLAKTVSVEE	LLQYVELQDP	300
VTRTQLRAMA	AKTVCPPHKV	ELEALLEKQA	YKEQVLAKRL	TMLELLEKYP	ACEMEFSEFI	360
ALLPSIRPRY	YSISSSPRVD	EKQASITVSV	VSGEAWSGYG	EYKGIASNYL	AELQEGDTIT	420
CFISTPQSEF	TLPKDPETPL	IMVGP GTGVA	PFRGFVQARK	QLKEQGQSLG	EAHLYFGCRS	480
PHEDYLYQEE	LENAQSEGII	TLHTAFSRMP	NQPKTYVQHV	MEQDGKKLIE	LLDKGAHFYI	540
CGDGSQMAPA	VEATLMKSYA	DVHQVSEADA	RLWLQQLLEK	GRYAKDVA		588

## SEQ ID NO:121

ccatcaaga

9

SEQ ID NO:122

PSR

3

**WHAT IS CLAIMED IS:**

1. A recombinant host comprising one or more of:
  - (a) a gene encoding an ent-kaurene oxidase (KO) polypeptide;
  - (b) a gene encoding a cytochrome P450 reductase (CPR) polypeptide; and/or
  - (c) a gene encoding an ent-kaurenoic **acid** hydroxylase (KAH) polypeptide;wherein at least one of the genes is a recombinant gene; and  
wherein the recombinant host is capable of producing a steviol glycoside precursor.
2. A recombinant host comprising:
  - (a) a gene encoding a geranylgeranyi diphosphate synthase (GGPPS) polypeptide;
  - (b) a gene encoding an ent-copalyl diphosphate synthase (CDPS) polypeptide;
  - (c) a gene encoding an ent-kaurene synthase (KS) polypeptide
  - (d) a gene encoding an ent-kaurene oxidase (KO) polypeptide;
  - (e) a gene encoding a cytochrome P450 reductase (CPR) polypeptide; and
  - (f) a gene encoding an ent-kaurenoic acid hydroxylase (KAH) polypeptide;wherein at least one of the **genes** is a recombinant gene; and  
wherein the recombinant host is capable of producing steviol.
3. The recombinant host of claims 1 or 2, wherein:
  - (a) the KO polypeptide comprises a KO polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:72 or SEQ ID NO:75; at least 65% identity to an amino acid sequence set forth in SEQ ID NO:54; at least 70% identity to an amino acid sequence set forth in SED ID NO: 70, SEQ ID NO:71, or

SEQ ID NO:79; at least 40% identity to an amino acid sequence set forth in SEQ ID NO:77; or at least 50% identity to an amino acid sequence set forth in SEQ ID NO:78;

- (b) the CPR polypeptide comprises a CPR polypeptide having at least 70% identity to an amino acid sequences set forth in SEQ ID NO:69, SEQ ID NO:74, SEQ ID NO:76, or SEQ ID NO:87; at least 80% identity to an amino acid sequence set forth in SEQ ID NO:73; at least 85% identity to an amino acid sequence set forth in SEQ ID NO:22; at least 65% identity to an amino acid sequence set forth in SEQ ID NO:28; or at least 50% identity to an amino acid sequence set forth in SEQ ID NO:98; and/or
- (c) the KAH polypeptide comprises a KAH polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:82; at least 50% identity to an amino acid sequence set forth in SEQ ID NO:91 ; or at least 60% identity to an amino acid sequence set forth in SEQ ID NO:68.

4 . A recombinant host comprising one or more of:

- (a) a gene encoding a KO polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:75;
- (b) a gene encoding a KAH polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:82; and/or
- (c) a gene encoding a CPR polypeptide having at least 50% identity to an amino acid sequence set forth in SEQ ID NO:98;

wherein at least one of the **genes** is a recombinant gene; and

wherein the recombinant host is capable of producing a steviol glycoside precursor.

5. A recombinant host comprising one or more of:

- (a) a gene encoding a KO polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:70;

- (b) a gene encoding a KAH polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:82; and/or
- (c) a gene encoding a CPR polypeptide having at least 50% identity to an amino acid sequence set forth in SEQ ID NO:98;

wherein at least one of the genes is a recombinant gene; and

wherein the recombinant host is capable of producing a steviol glycoside precursor.

- 6. The recombinant host of claim 4 or 5, wherein the host further comprises a gene encoding a KO polypeptide having at least 65% identity to an amino acid sequence set forth in SEQ ID NO:54.
- 7. The recombinant host of any one of claims 4-6, wherein the host further comprises a gene encoding a KAH polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:68.
- 8. The recombinant host of any one of claims 4-7, wherein the host further comprises a gene encoding a KO polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:79.
- 9. The recombinant host of any one of claims 1 or 3-8, wherein the host further comprises one or more of:
  - (a) a gene encoding a geranylgeranyl diphosphate synthase (GGPPS) polypeptide;
  - (b) a gene encoding an ent-copalyl diphosphate synthase (CDPS) polypeptide; and/or
  - (c) a gene encoding an ent-kaurene synthase (KS) polypeptide;

wherein at least one of the genes is a recombinant gene; and



wherein the recombinant host is capable of producing a steviol glycoside precursor.

10. The recombinant host of claim 9, wherein:
  - (a) the GGPPS polypeptide comprises a polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:49;
  - (b) the CDPS polypeptide comprises a polypeptide having at least 70% identity to an amino acid sequence set forth in SEQ ID NO:37; and/or
  - (c) the KS polypeptide comprises a polypeptide having at least 40% identity to an amino acid sequence set forth in SEQ ID NO:6.
11. The recombinant host of claims 1-10, wherein the host further comprises a gene encoding an endoplasmic reticulum membrane polypeptide.
12. The recombinant host of claim 11, wherein the endoplasmic reticulum membrane polypeptide comprises an Inheritance of cortical ER protein 2 (ICE2) polypeptide having at least 50% identity to the amino acid sequence set forth in SEQ ID NO:114.
13. The recombinant host of any one of claim 1-10, wherein the KO polypeptide is a fusion construct.
14. The recombinant host of claim 13, wherein the fusion construct comprises a polypeptide having at least 60% identity to an amino acid sequence set forth in SEQ ID NO:118 or SEQ ID NO:120.
15. The recombinant host of claim 13 or claim 14, wherein the fusion construct has at least 50% identity to an amino acid sequence set forth in SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:106, SEQ ID NO:108, SEQ ID NO:110, or SEQ ID NO:112.

16. The recombinant host of any one of claims 1-15, wherein the host further comprises one or more of:

- (a) a gene encoding a UGT85C polypeptide;
- (b) a gene encoding a UGT76G polypeptide;
- (c) a gene encoding a UGT74G1 polypeptide;
- (d) a gene encoding a UGT91 D2 functional homolog polypeptide; and/or
- (e) a gene encoding an EUGT11 polypeptide;

wherein at least one of the genes is a recombinant gene; and

wherein the host is capable of producing a steviol glycoside.

17. The recombinant host of claim 16, wherein:

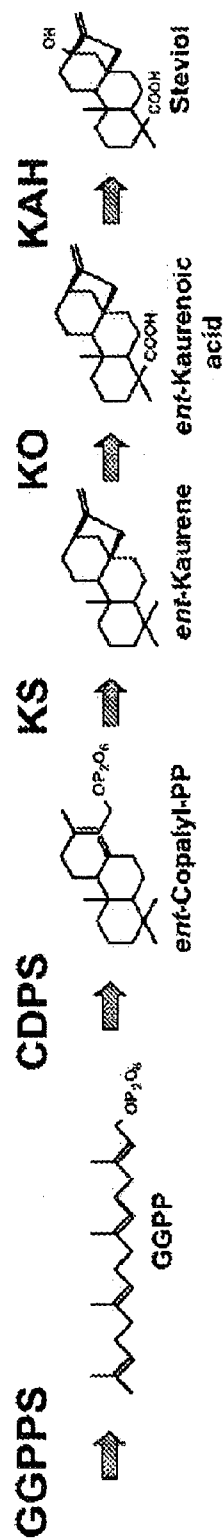
- (a) the UGT85C2 polypeptide comprises a polypeptide having at least 55% identity to an amino acid sequence set forth in SEQ ID NO:30;
- (b) the UGT76G1 polypeptide comprises a polypeptide having at least 50% identity to an amino acid sequence set forth in SEQ ID NO:83;
- (c) the UGT74G1 polypeptide comprises a polypeptide having at least 55% identity to an amino acid sequence set forth in SEQ ID NO:29;
- (d) the UGT91 D2 functional homolog polypeptide comprises a UGT91D2 polypeptide having 90% or greater identity to the amino acid sequence set forth in SEQ ID NO:84 or a UGT91D2e-b polypeptide having 90% or greater identity to the amino acid sequence set forth in SEQ ID NO:88; and/or
- (e) the EUGT11 polypeptide comprises a polypeptide having at least 65% identity to an amino acid sequence set forth in SEQ ID NO:86.

18. The recombinant host of any one of claims 1-17, wherein the recombinant host comprises a plant cell, a mammalian cell, an insect cell, a fungal cell, or a bacterial cell.

- 1<sub>19</sub>. The recombinant host of claim 18, wherein the bacterial cell comprises *Escherichia* bacteria cells, *Lactobacillus* bacteria cells, *Lactococcus* bacteria cells, *Cornebacterium* bacteria cells, *Acetobacter* bacteria cells, *Acinetobacter* bacteria cells, or *Pseudomonas* bacterial cells.
20. The recombinant host of claim 18, wherein the fungal cell comprises a yeast cell.
- 2 1<sub>1</sub>. The recombinant host of claim 20, wherein the yeast cell is a cell from *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Candida glabrata*, *Ashbya gossypii*, *Cyberlindnera jadinii*, *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha*, *Candida boidinii*, *Arxula adeninivorans*, *Xanthophyllomyces dendrorhous*, or *Candida albicans* species.
22. The recombinant host of claim 21, wherein the yeast cell is a *Saccharomycete*.
23. The recombinant host of claim 22, wherein the yeast cell is a cell from the *Saccharomyces cerevisiae* species.
24. A method of producing a steviol glycoside or a steviol glycoside precursor, comprising:
- (a) growing the recombinant host of any one of claims 1-23 in a culture medium, under conditions in which any of the genes disclosed in any one of claims 1-23 are expressed;  
wherein the **steviol** glycoside or the steviol glycoside precursor is synthesized by said host; and/or
  - (b) optionally quantifying the steviol glycoside or the steviol glycoside precursor; and/or
  - (c) optionally isolating the steviol glycoside or the steviol glycoside precursor.

25. The method of claim 24, wherein the steviol glycoside comprises steviol-1 3-O-glucoside (13-SMG), steviol-1 ,2-bioside, steviol-1,3-bioside, steviol-1 9-O-glucoside (19-SMG), stevioside, 1,3-stevioside, rubusoside, Rebaudioside A (RebA), Rebaudioside B (RebB), Rebaudioside C (RebC), Rebaudioside D (RebD), Rebaudioside E (RebE), Rebaudioside F (RebF), Rebaudioside M (RebM), Rebaudioside Q (RebQ), Rebaudioside I (RebI), dulcoside A, di-glycosylated steviol, tri-glycosylated steviol, tetra-glycosylated steviol, penta-glycosylated steviol, hexa-glycosylated steviol, hepta-glycosylated steviol, or isomers thereof.
26. The steviol glycoside or the steviol glycoside precursor produced by the recombinant host of any one of claims 1-23 or the method of claim 24 or claim 25, wherein the steviol glycoside or steviol glycoside precursor accumulates to a detectable concentration when cultured under said conditions.
27. A steviol glycoside composition produced by the host of any one of claims 1-23 or the method of claim 24 or claim 25, wherein the composition has an undetectable concentration of stevia plant-derived contaminants.
28. A steviol glycoside composition produced by the host of any one of claims 1-23 or the method of claim 24 or claim 25, wherein the composition has a steviol glycoside composition enriched for RebD or RebM relative to the steviol glycoside composition of a wild-type Stevia plant.

Figure 1



## Figure 2

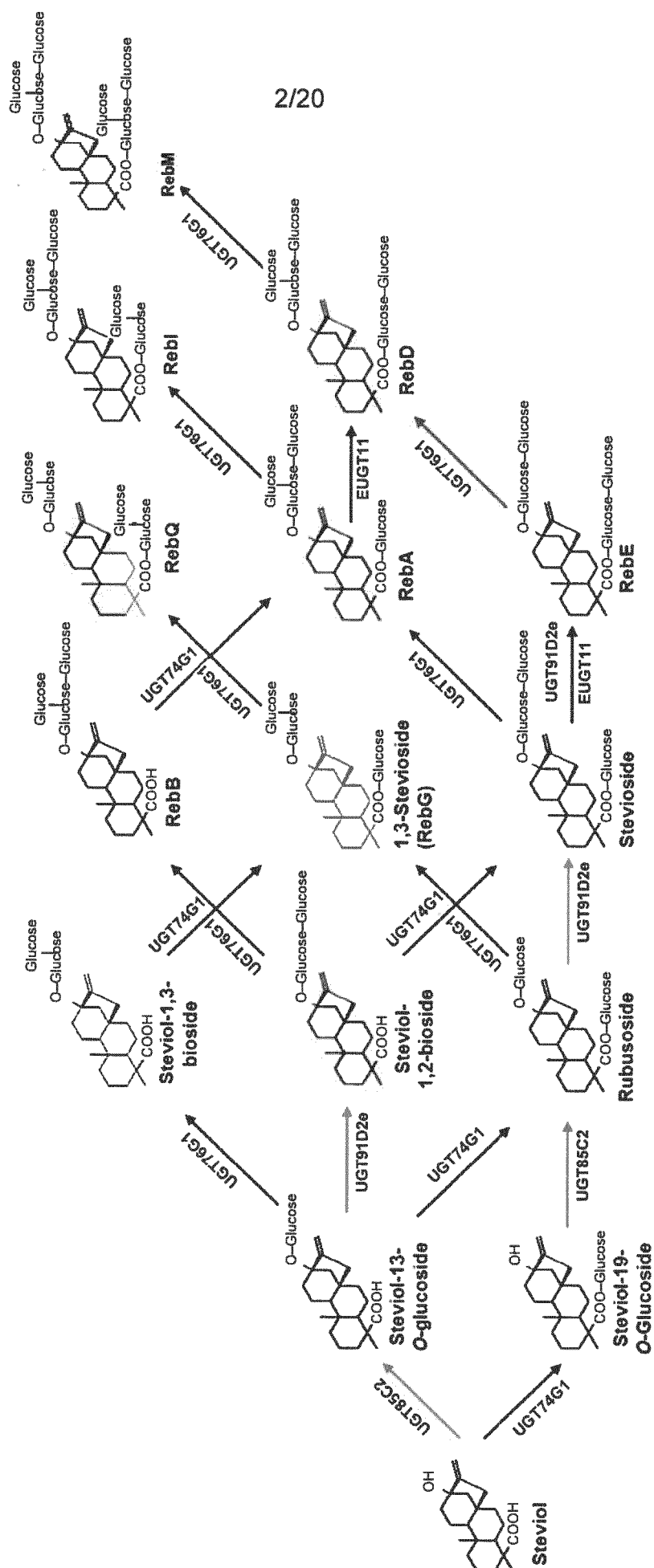


Figure 3

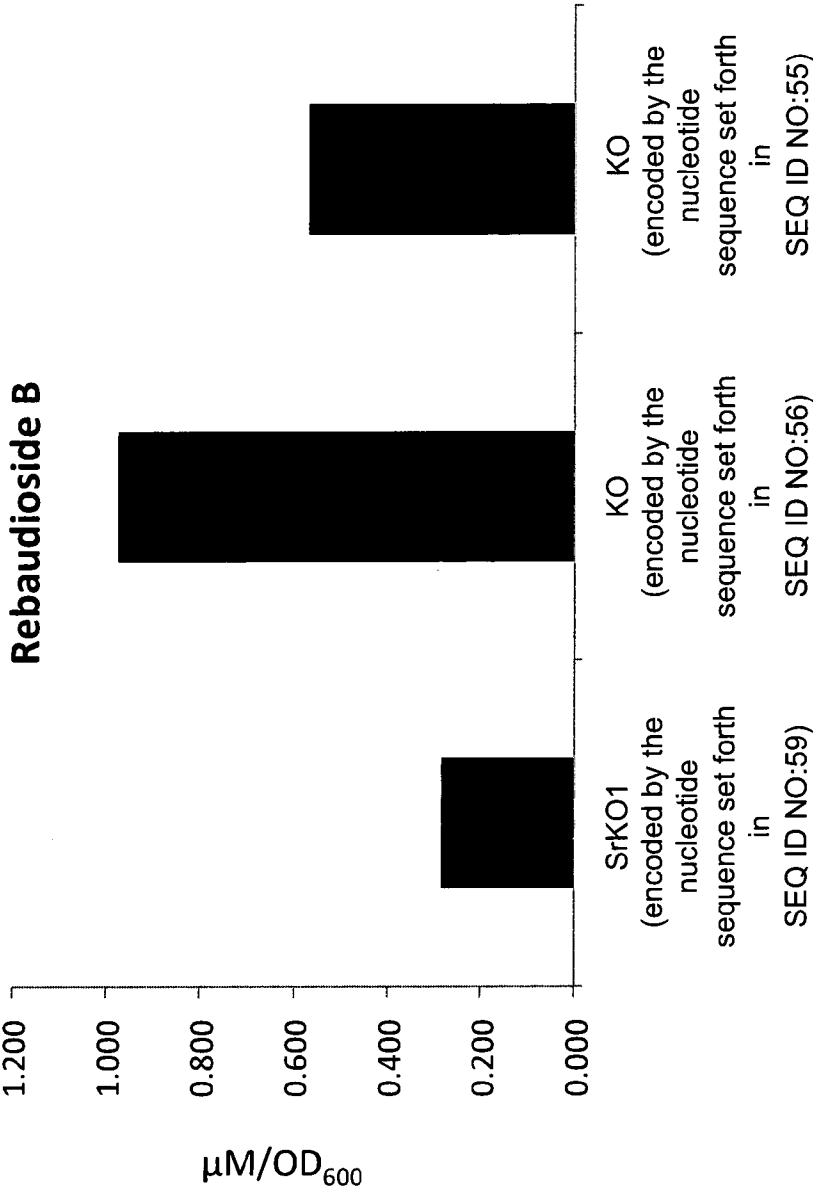
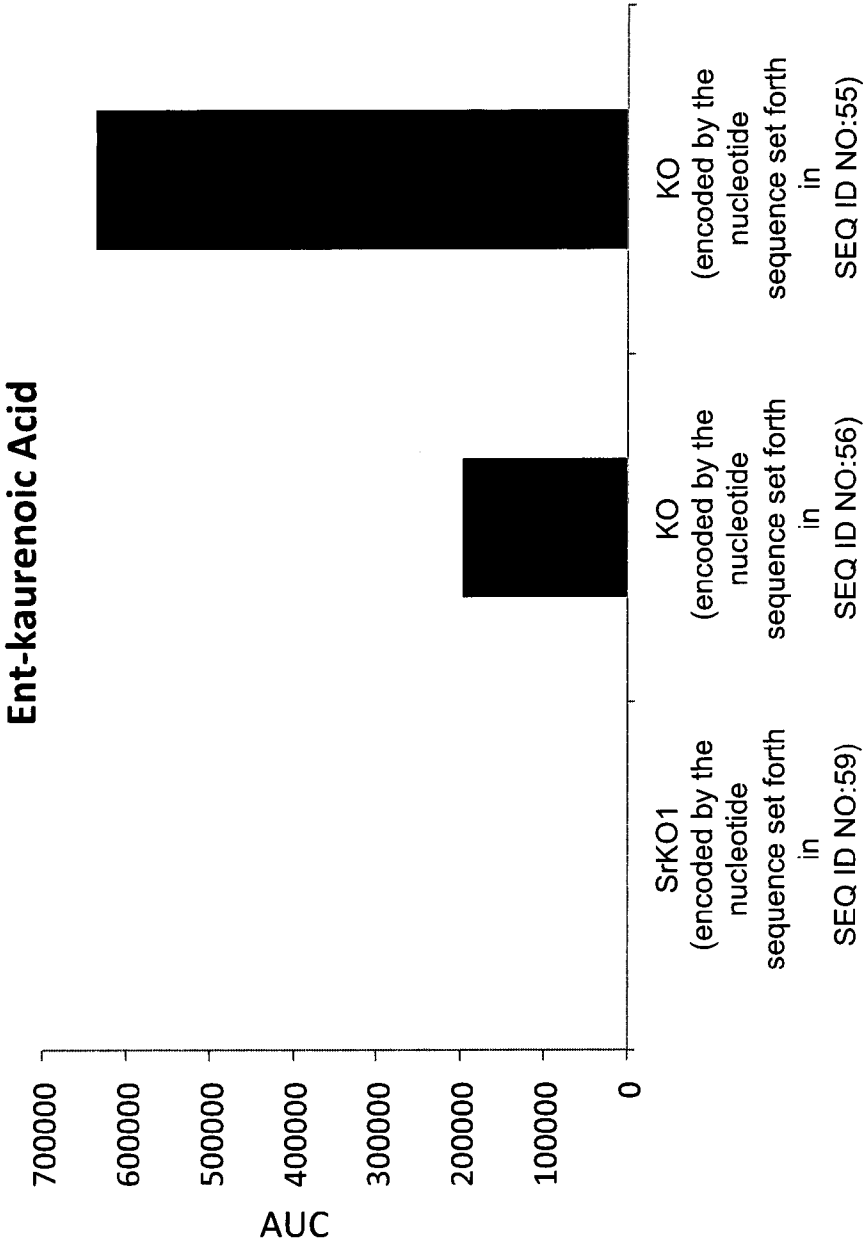


Figure 4





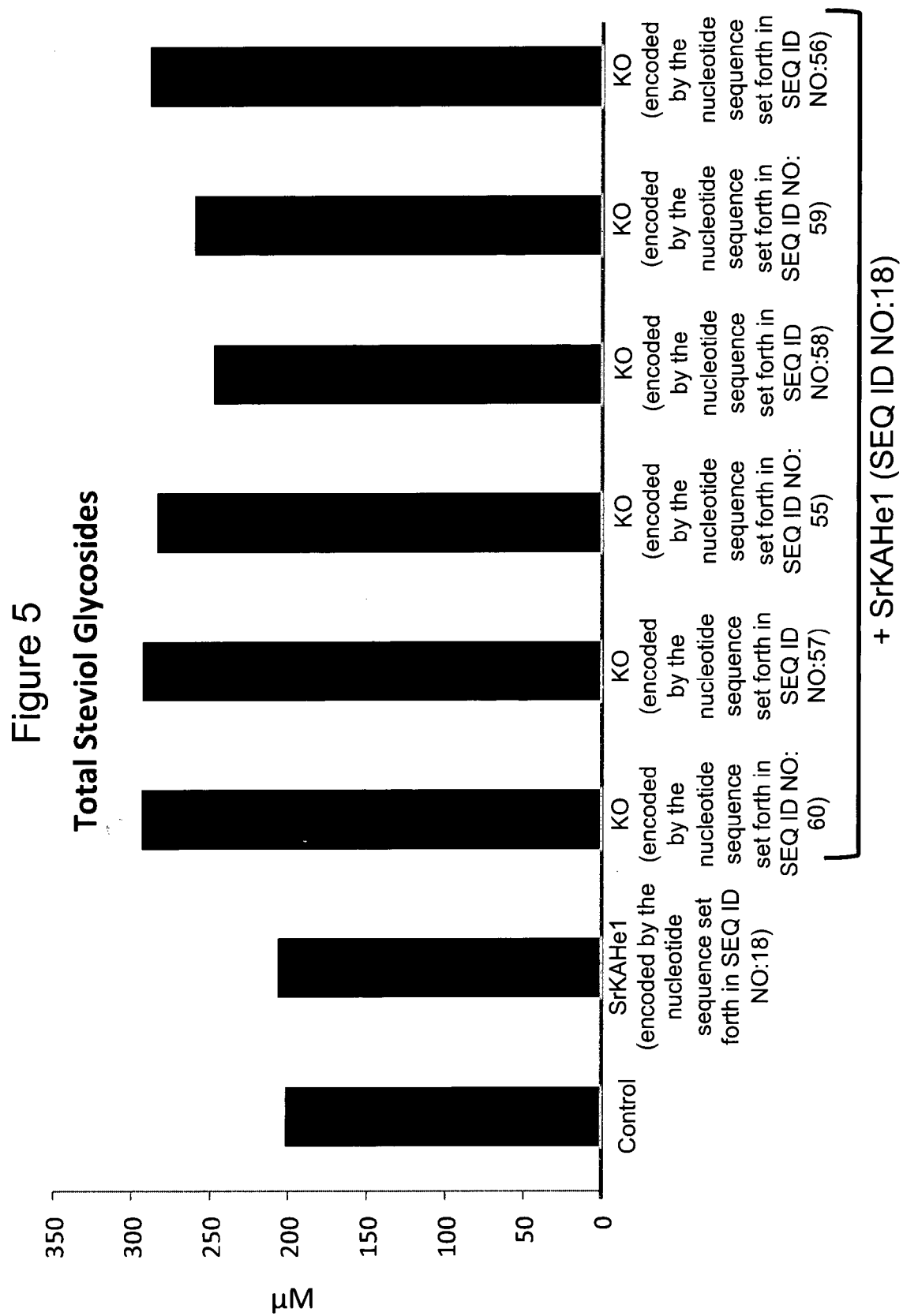


Figure 6

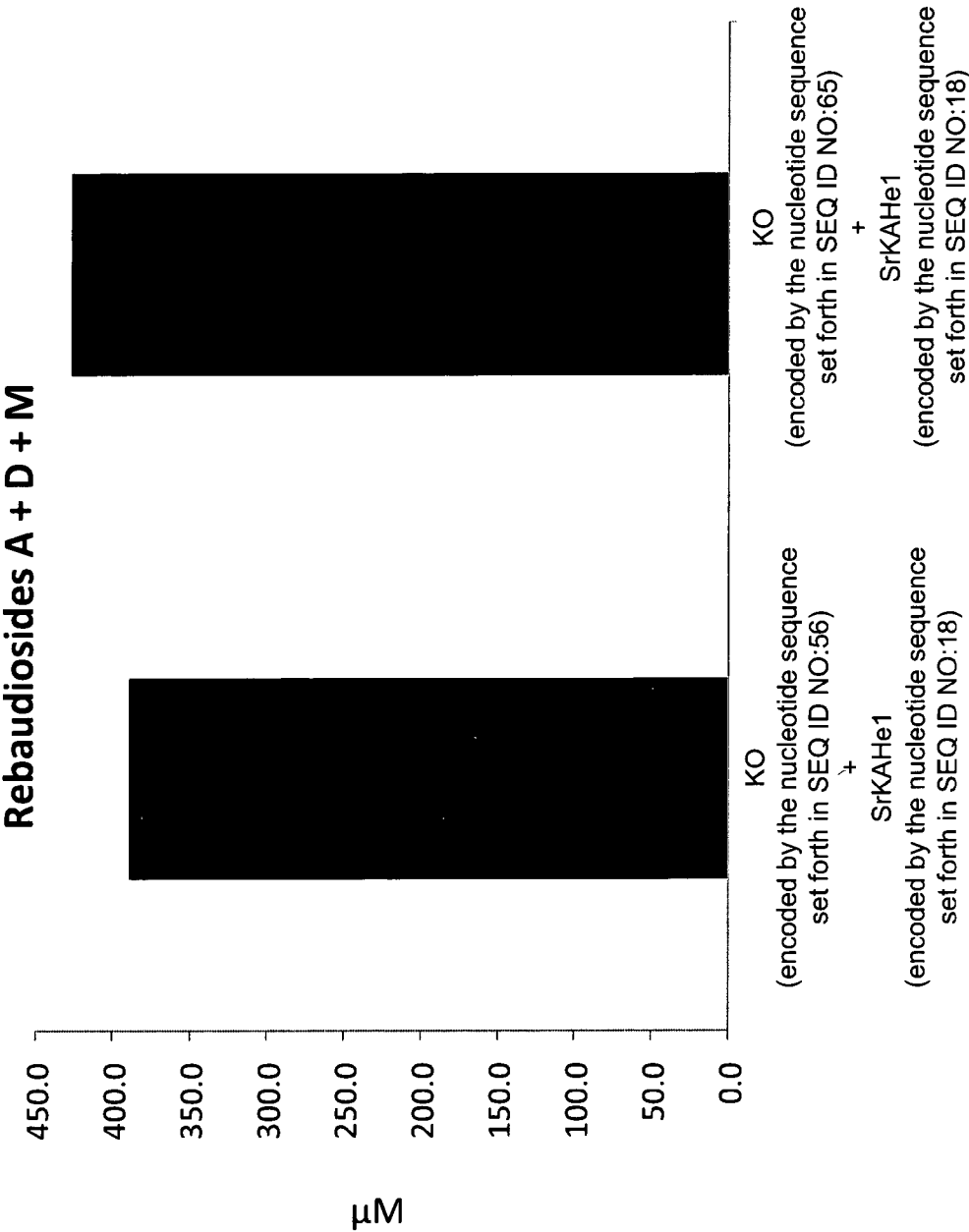
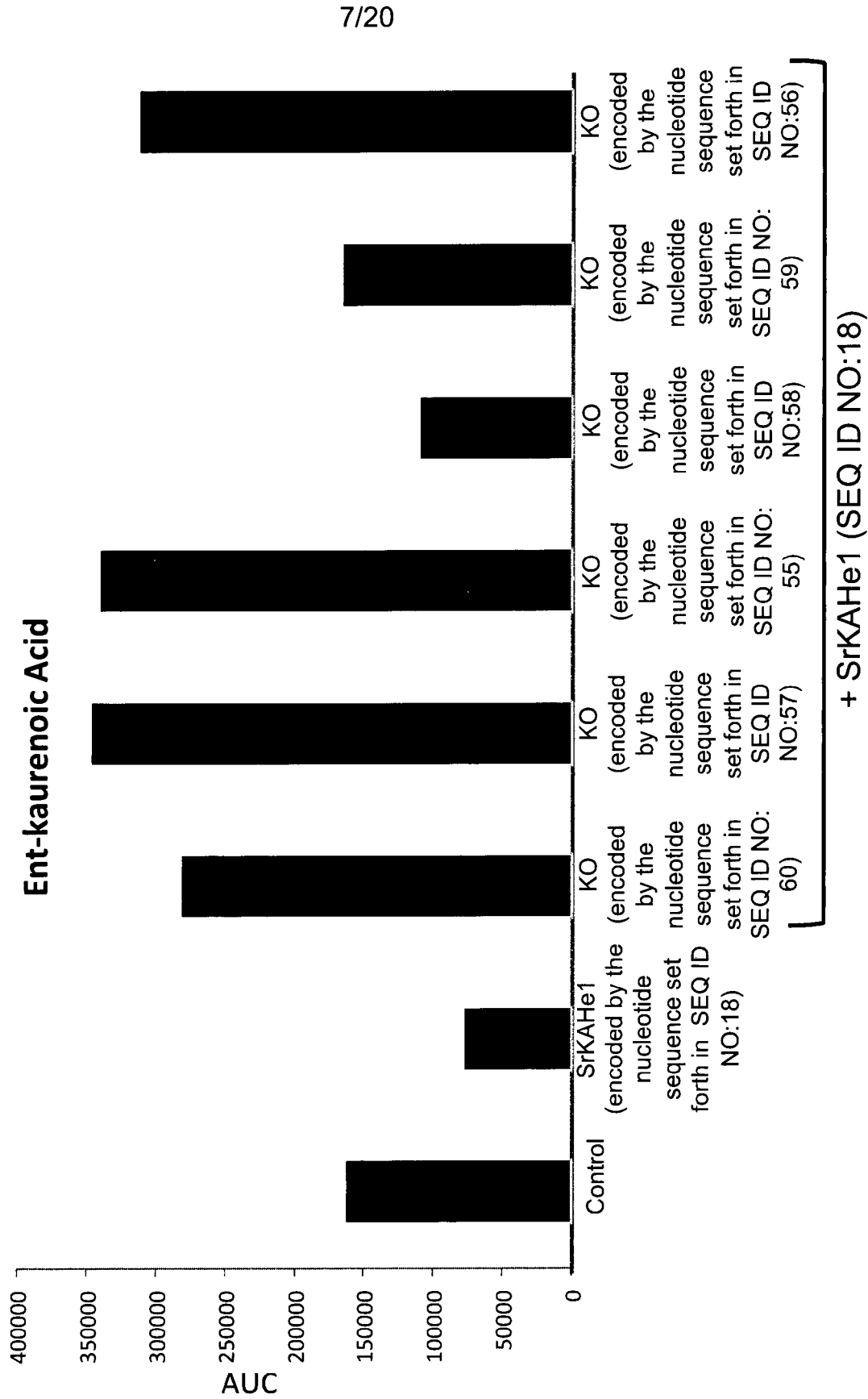
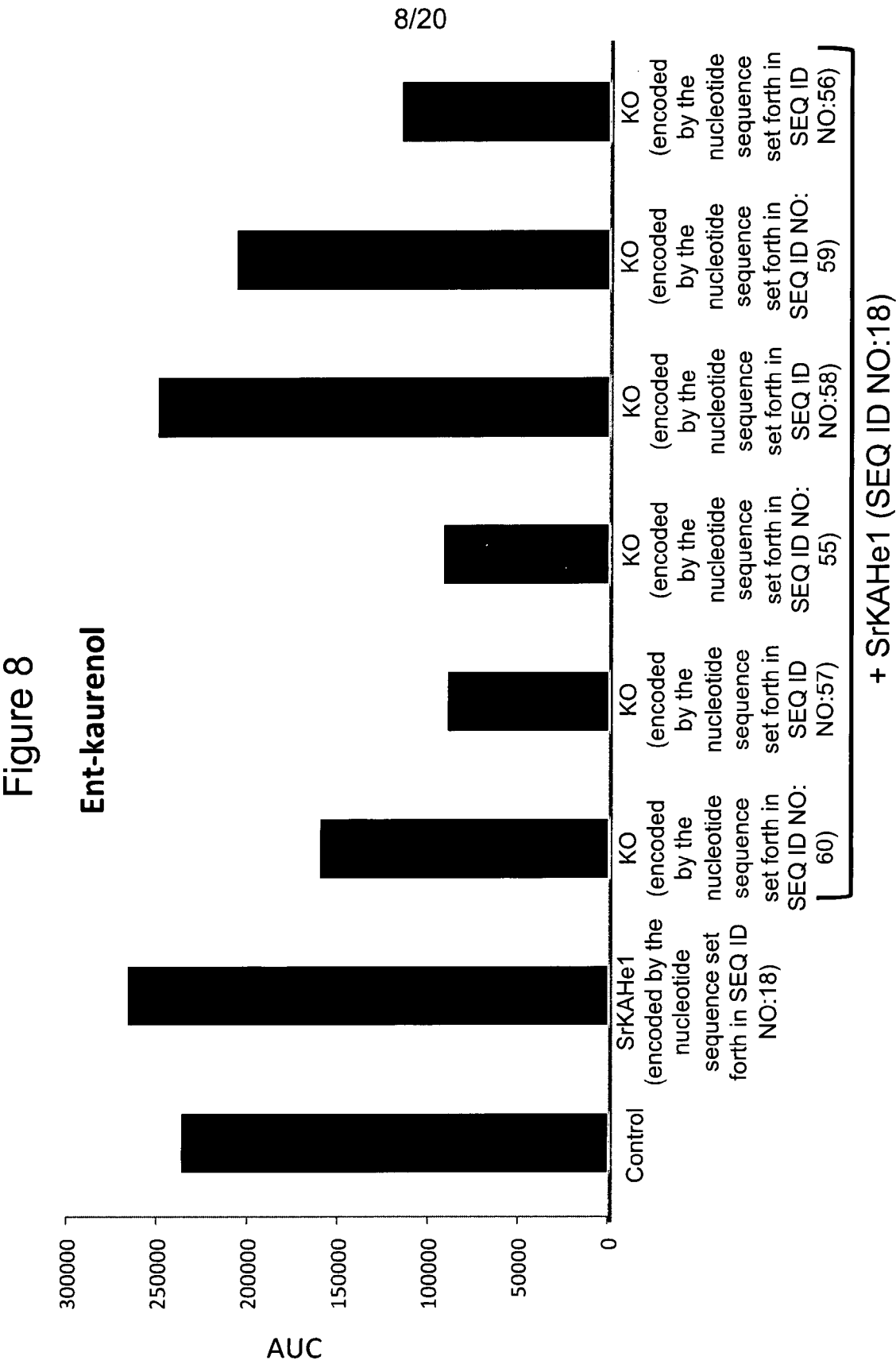


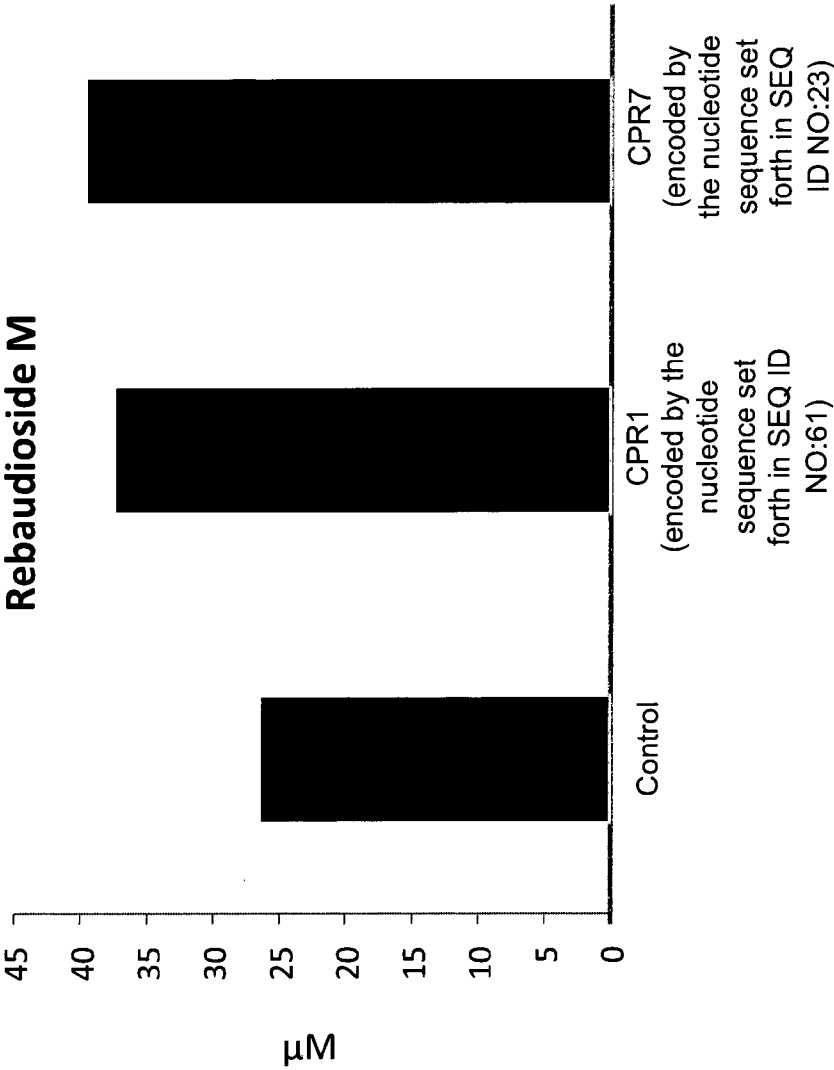
Figure 7





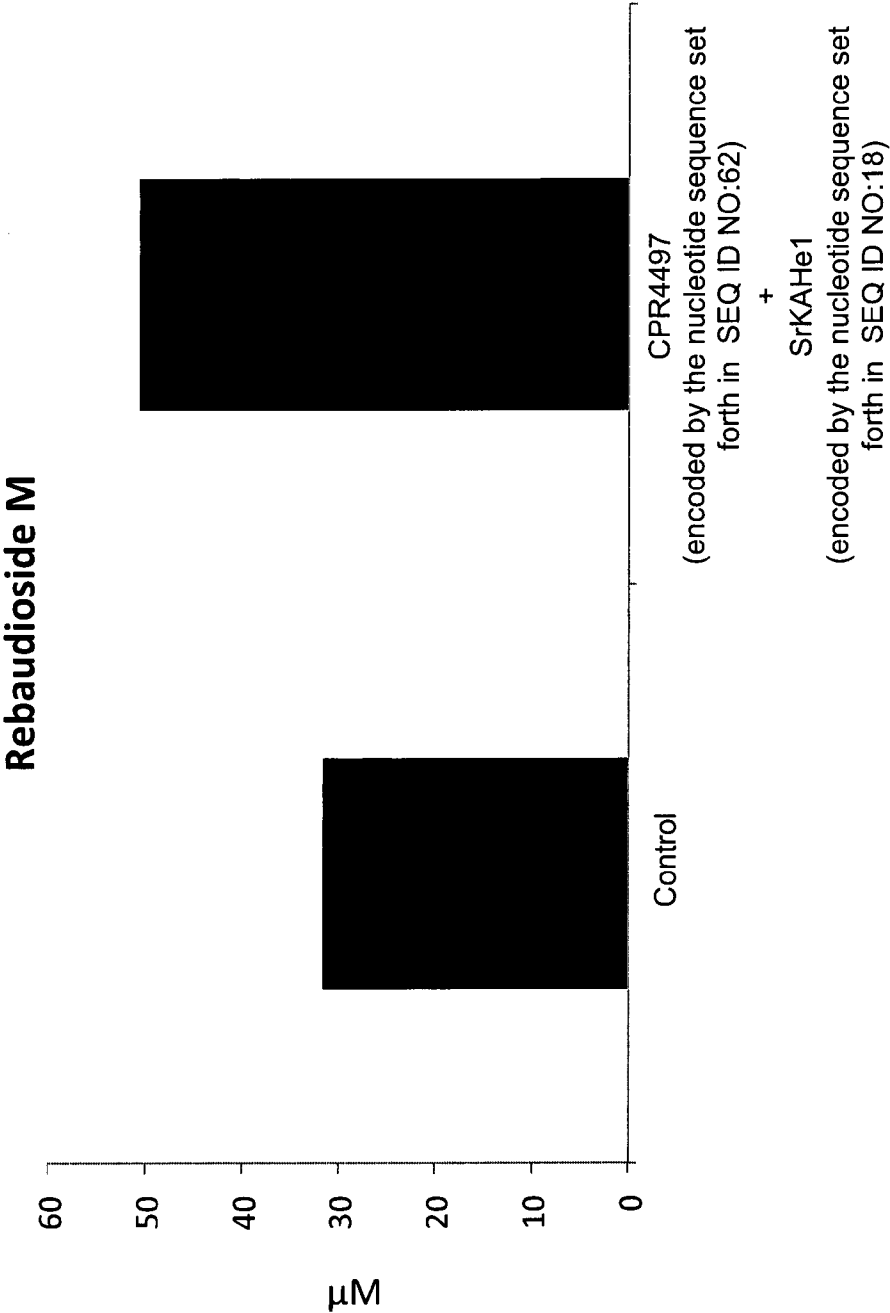
9/20

Figure 9



10/20

Figure 10



11/20

Figure 11A

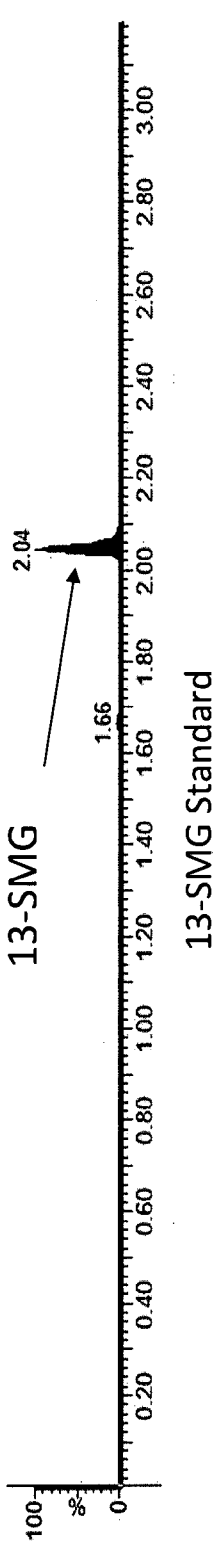
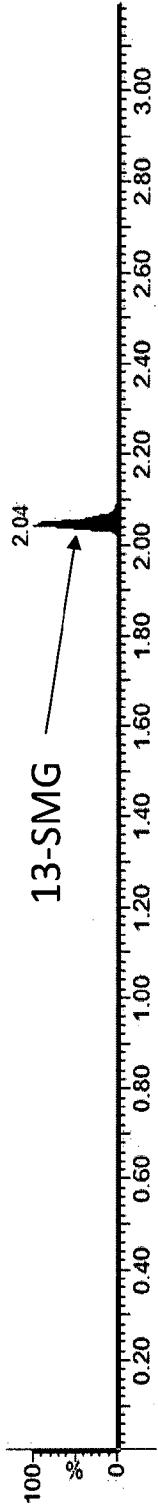


Figure 11B



*S. cerevisiae* expressing KAH encoded by nucleotide sequence set forth in SEQ ID NO:80

Figure 12

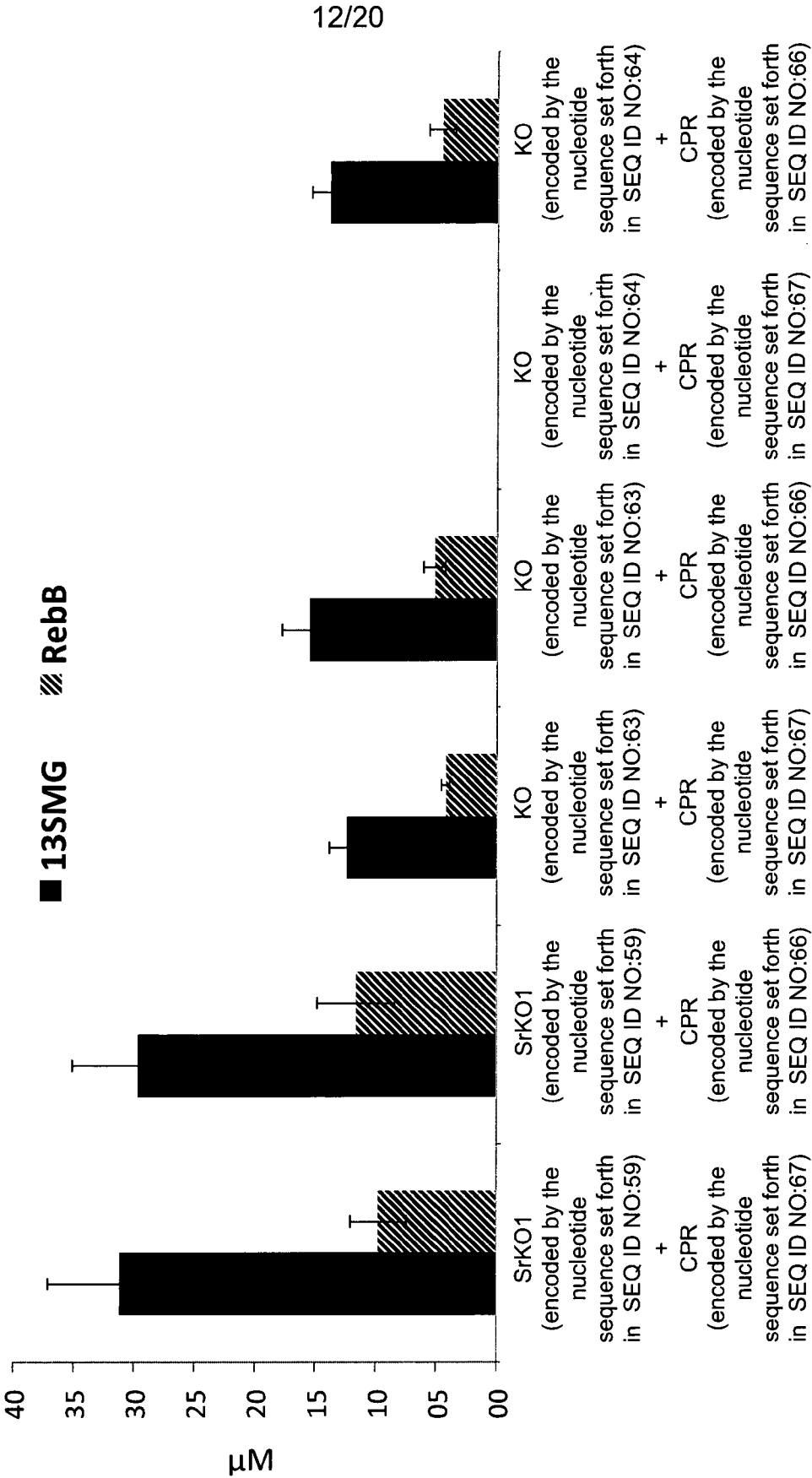




Figure 13

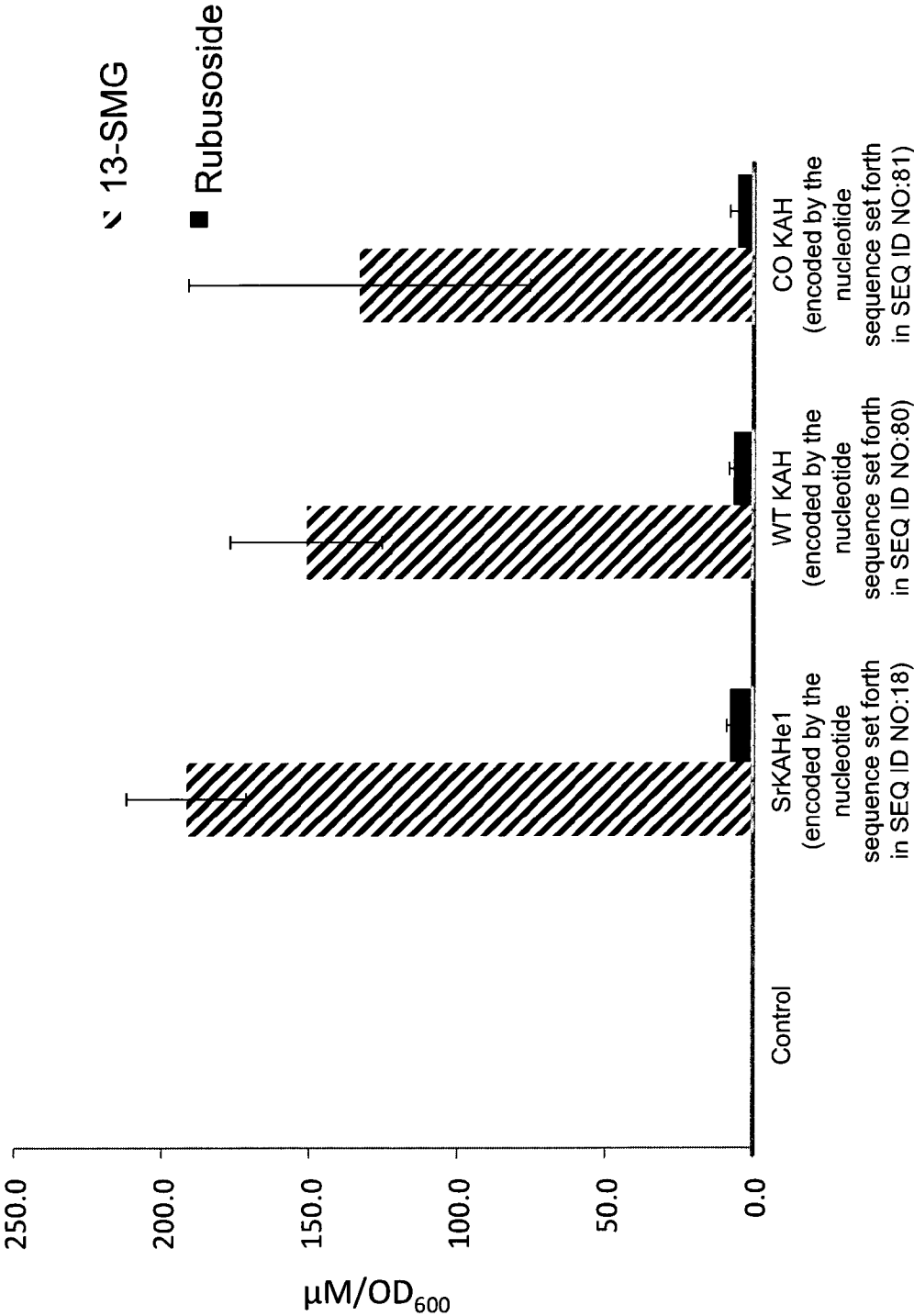


Figure 14

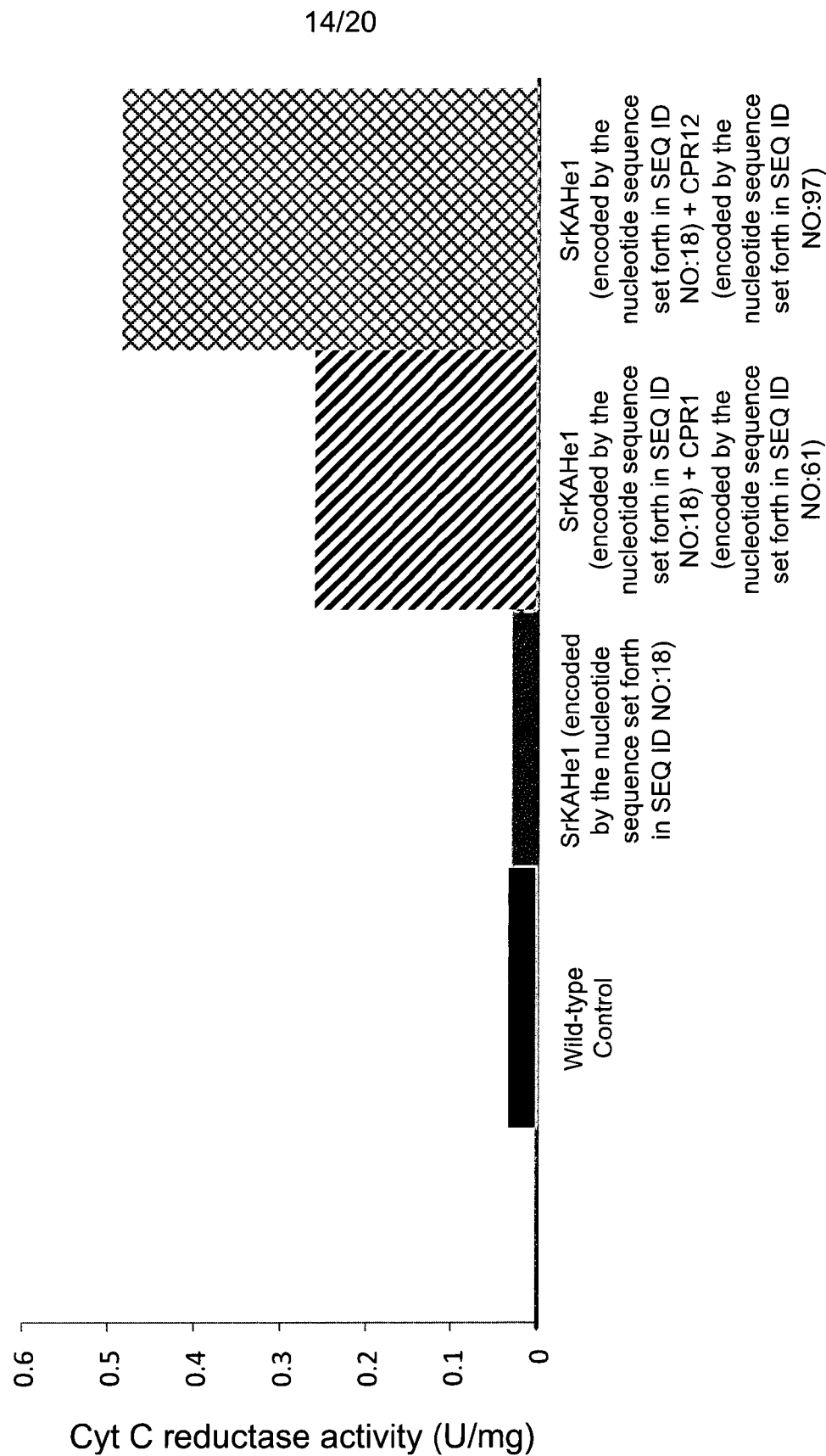


Figure 15A

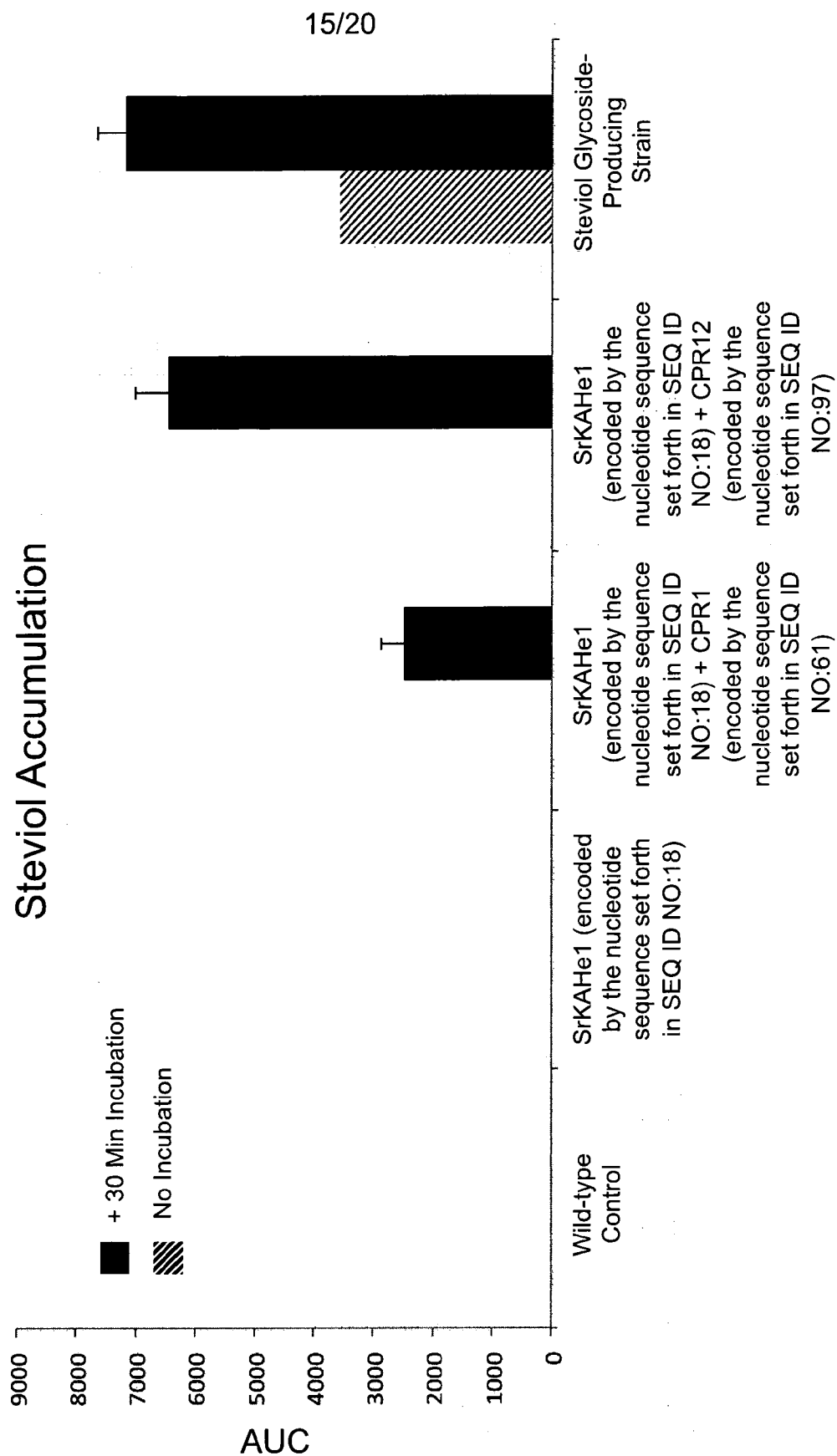


Figure 15B

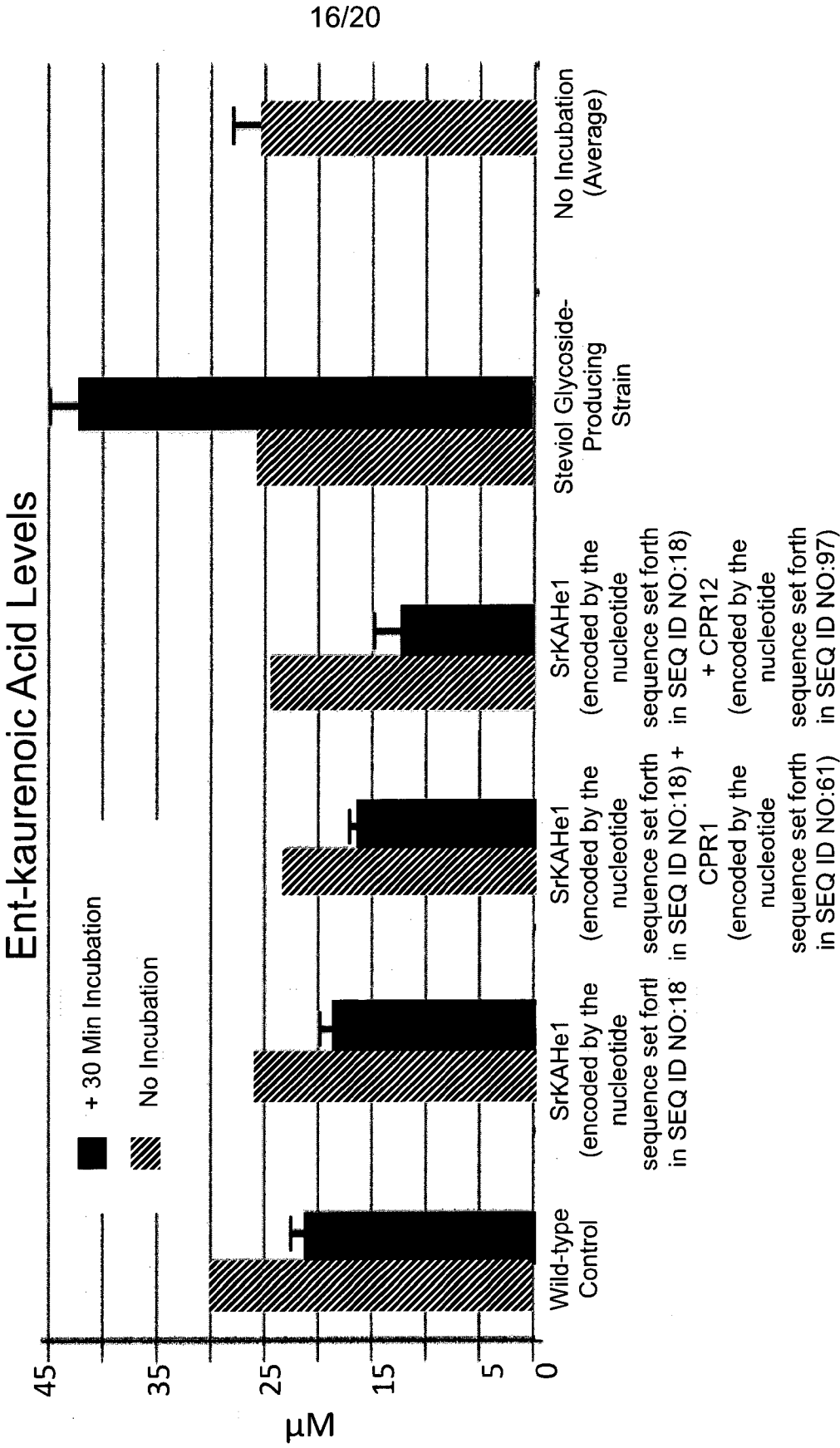


Figure 16A

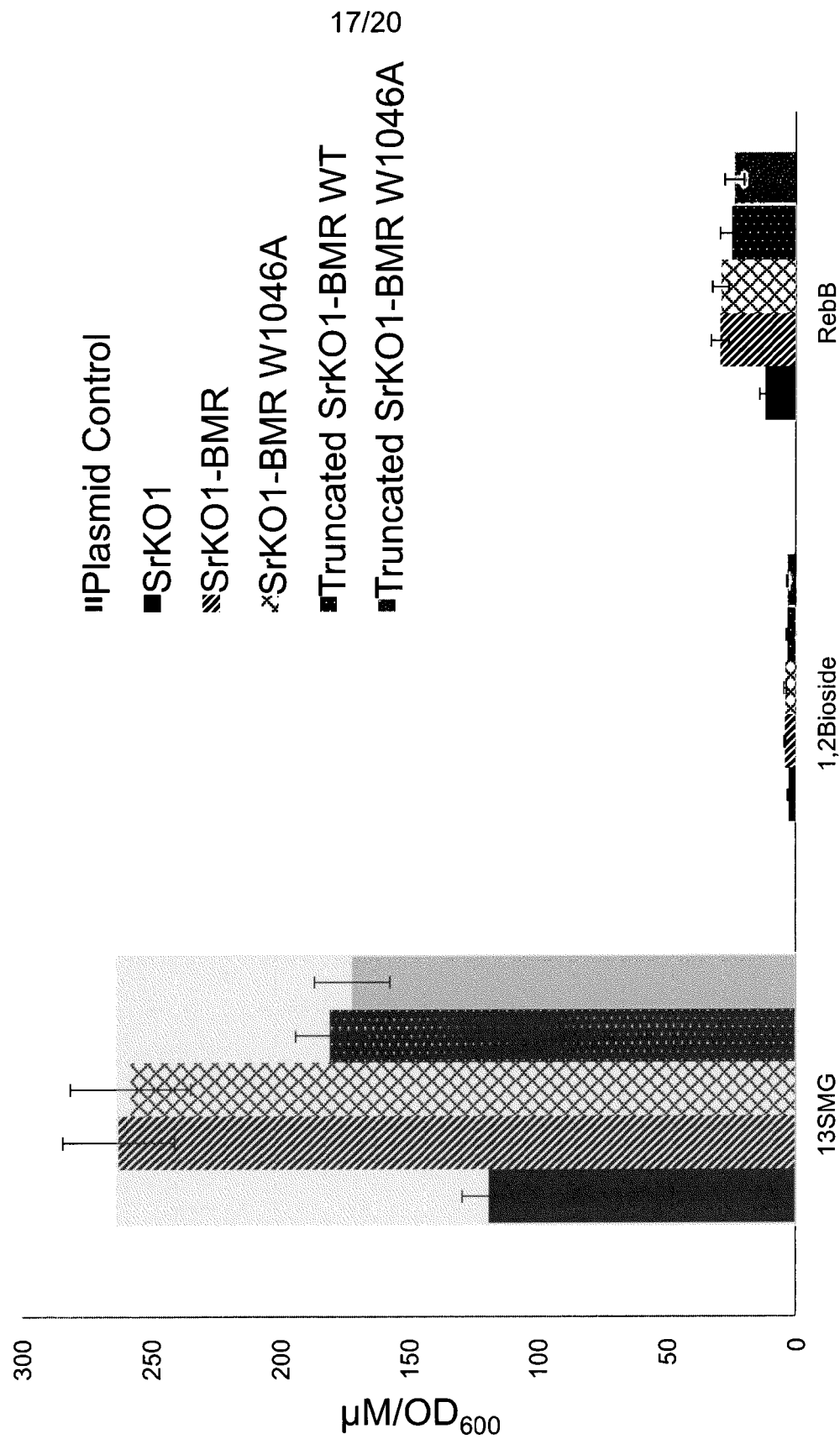


Figure 16B

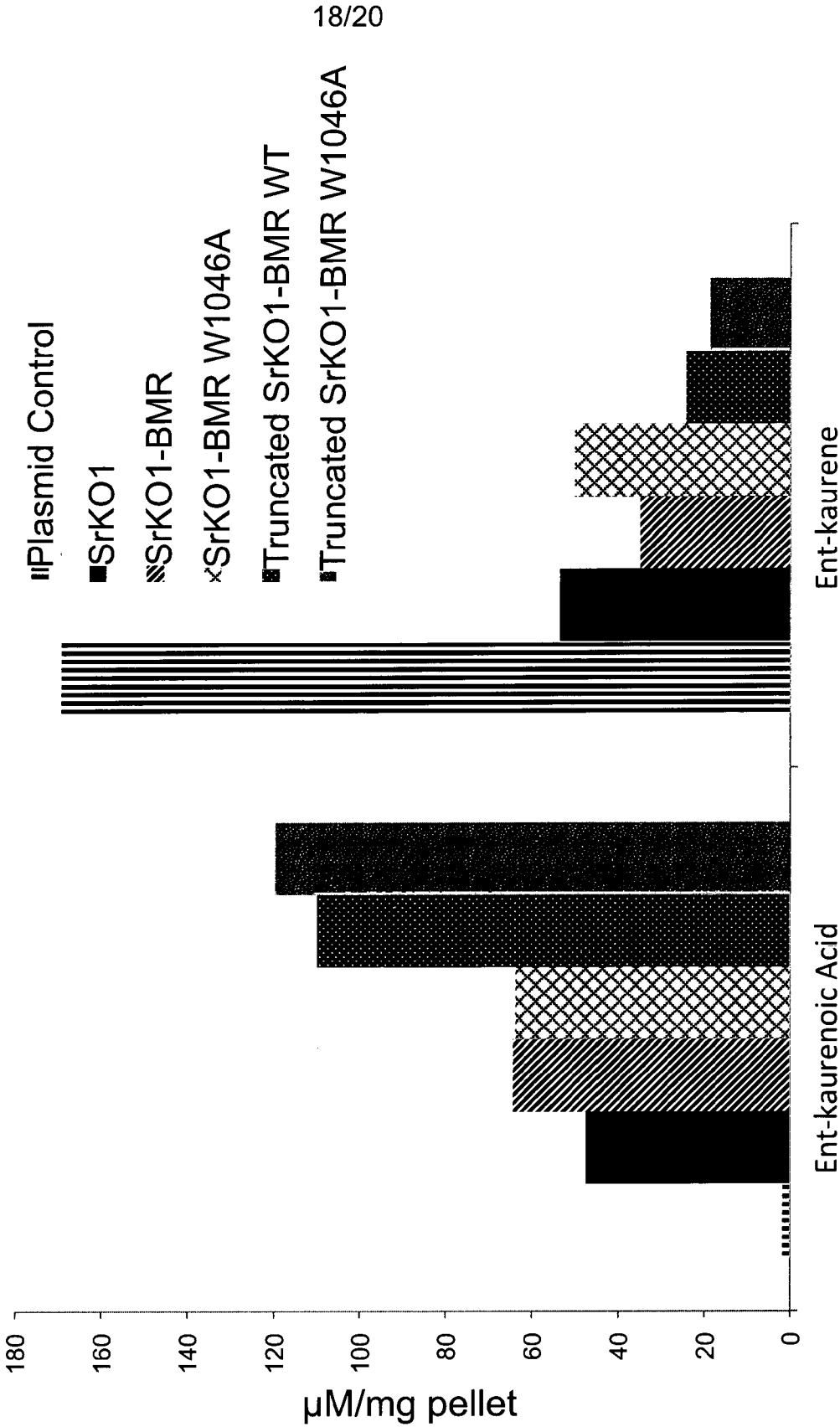


Figure 16C

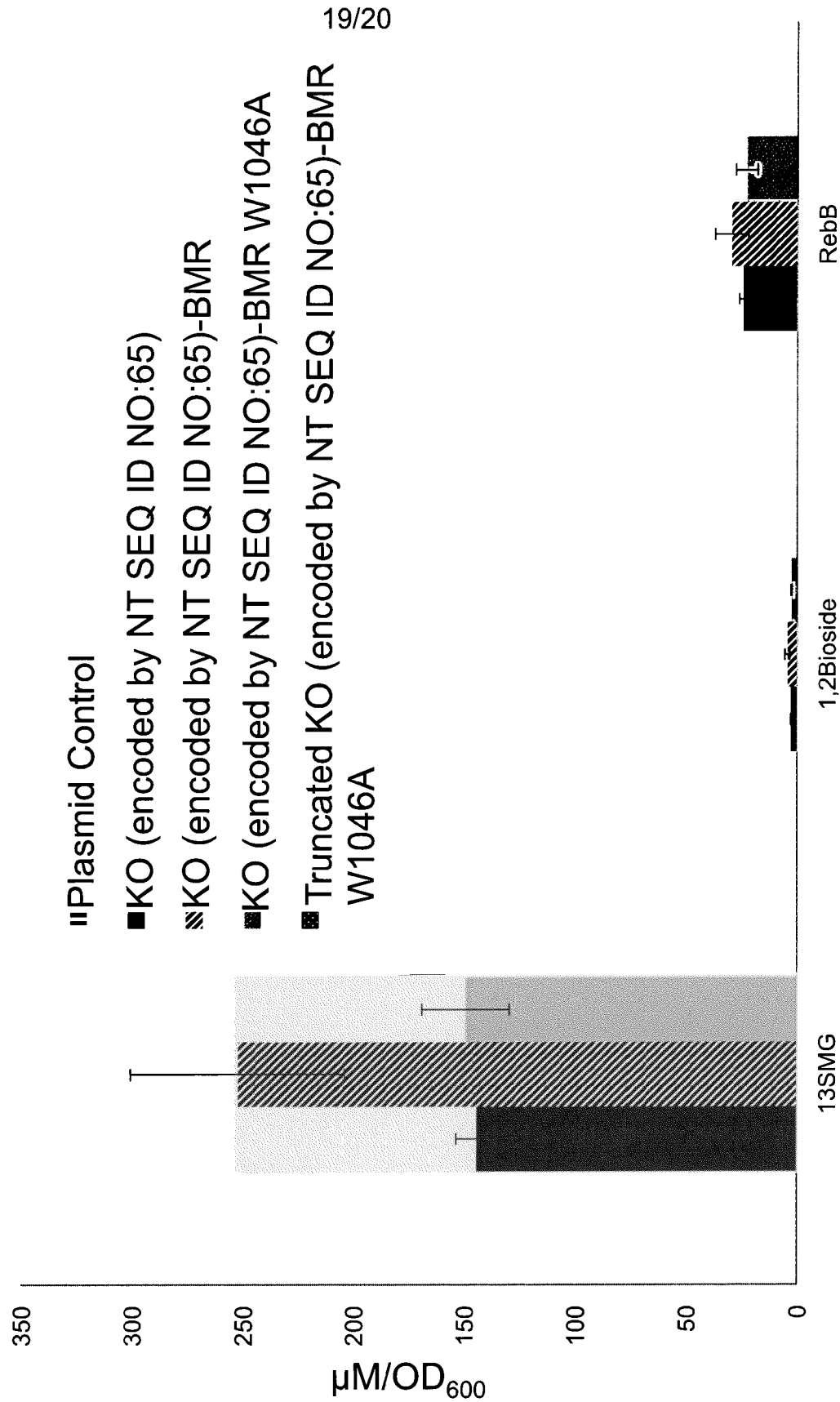


Figure 16D

